MAPPING URBAN CHANGE USING STREET VIEW IMAGERY AND COMPUTER VISION: CASE OF KARAKÖY

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

ARCHITECTURE

By

Elifnaz Kancan

August 2025

MAPPING URBAN CHANGE USING STREET VIEW IMAGERY AND COMPUTER VISION: CASE OF KARAKÖY

By Elifnaz Kancan

August 2025

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Burcu Şenyapılı Özcan(Advisor)

Yiğit Acar(Co-Advisor)

Pelin Yoncacı Arslan

Onur Tümtürk

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

MAPPING URBAN CHANGE USING STREET VIEW IMAGERY AND COMPUTER VISION: CASE OF KARAKÖY

Elifnaz Kancan
M.S. in Architecture
Advisor: Burcu Şenyapılı Özcan
Co-Advisor: Yiğit Acar
August 2025

Traditional urban studies have relied on fieldwork and archival research. Yet, recent advances in AI and remote sensing provide scalable methods to monitor urban transformation. By comparing historical and recent Street View imagery, computer vision enables the systematic detection of spatial modifications. However, if not critically linked to socio-historical frameworks, such outputs risk becoming decontextualized. This study introduces the concepts of pings and signals within a mapping-based model that combines computational detection with contextual interpretation. Micro-scale interventions, such as demolitions, sign changes, or function shifts, are treated as pings. When accumulated over time, these pings evolve into signals that indicate broader shifts in land use and socio-economic dynamics. Karaköy, Istanbul, is examined as a site where redevelopment and the growth of creative industries have reshaped the urban fabric. These processes have altered the area's identity, accessibility, and spatial practices. The findings suggest that Google Street View, when used in conjunction with diachronic models, can effectively support spatial analysis by capturing patterns of change. Nevertheless, this approach has several limitations. These include data biases, uneven temporal coverage, and limited qualitative integration, all of which require careful and critical interpretation. Despite these challenges, street-level imagery remains a valuable source. This article demonstrates that the integration of computational and experiential data enables richer interpretations of urban change. In this context, mapping emerges as a relational practice that links physical traces to broader contextual narratives.

Keywords: Urban change, computer vision, remote sensing, mapping.

ÖZET

SOKAK GÖRÜNTÜLERİ VE BİLGİSAYARLI GÖRÜ İLE KENTSEL DEĞİŞİMİN HARİTALANDIRILMASI: KARAKÖY ÖRNEĞİ

Elifnaz Kancan
Mimarlık, Yüksek Lisans
Tez Danışmanı: Burcu Şenyapılı Özcan
Ėş Danışman: Yiğit Acar
Ağustos 2025

Geleneksel kent çalışmaları genellikle saha araştırmaları ve arşiv verilerine dayanmıştır; ancak yapay zekâ ve uzaktan algılamadaki son gelişmeler, kentsel dönüşümün izlenmesine yönelik ölçeklenebilir yöntemler sunmaktadır. Tarihsel ve güncel Street View görüntülerinin karşılaştırılmasıyla, bilgisayarlı görü mekânsal değişimlerin sistematik olarak tespit edilmesini sağlar; ancak bu yöntemler, sosyo-tarihsel çerçevelerle eleştirel biçimde ilişkilendirilmediğinde bağlamdan kopuk çıktılar üretme riski taşır. Bu çalışma, haritalama temelli bir model içinde "ping" ve "sinyal" kavramlarını tanıtarak hesaplamalı tespit ile bağlamsal yorumu birleştirmektedir. Yıkım, tabela değişikliği veya işlevsel dönüşüm gibi mikro ölçekli müdahaleler ping olarak değerlendirilmekte; bu ping'ler zaman içinde bir araya gelerek, arazi kullanımı ve sosyo-ekonomik dinamiklerdeki daha geniş değişimlere işaret eden sinyalleri oluşturmaktadır. İstanbul Karaköy örneği, yeniden gelişim ve yaratıcı endüstrilerin yayılımı ile kentsel dokunun yeniden şekillenişini; kimliğin, erişimin ve mekânsal pratiklerin dönüşümünü ortaya koymaktadır. Bulgular, Google Street View verisinin zamansal analiz modelleriyle birleştirildiğinde mekânsal analiz için etkili bir araç olabileceğini göstermektedir. Tüm potansiyeline rağmen bu yaklaşım; veri yanlılığı, düzensiz zamansal kapsama ve sınırlı nitel bağlam gibi sınırlılıkları da barındırmakta ve eleştirel bir yorum gerektirmektedir. Yine de sokak düzeyindeki görüntüler değerli bir kaynak olmayı sürdürmektedir. Bu çalışma, hesaplamalı ve deneyimsel verilerin kentsel dönüşümü nasıl daha katmanlı biçimde anlamaya katkı sunduğunu ve haritalamanın fiziksel izleri bağlamsal anlatılarla ilişkilendiren bir eylem olarak nasıl işlediğini ortaya koymaktadır.

Anahtar sözcükler: kentsel değişimler, bilgisayarlı görü, uzaktan algılama, haritalama.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to Dr.Yiğit Acar, who has been my advisor since my undergraduate years and served as the co-advisor of this thesis. I had the chance to learn from Dr. Acar through his academic guidance and through the experience of working closely with him in ARCH 401 and ARCH 402, where I served as a teaching assistant. These experiences allowed me to better understand professional ethics, educational responsibility, and the values of academic integrity. Dr. Acar supports the academic and personal development of many students beyond myself. He encourages the exploration of alternative design ideas, inspires long-term thinking, and is always attentive to new opportunities that could benefit his students. The Department of Architecture at Bilkent is truly fortunate to have him. I am also very thankful for his eagerness to learn new things together with me and for supporting me in this regard throughout the thesis process.

I am also grateful to Assoc. Prof. Dr. Burcu Şenyapılı Özcan, who served as the advisor of this thesis. Throughout the research process, her guidance and constructive feedback helped me stay on the right track. Through her graduate-level courses, particularly Architectural Readings, I was able to discover new dimensions of architectural thinking and refine my theoretical perspective. I thank her for her supervision, as well as her thoughtful comments and revisions that contributed to the quality of this work.

I would like to sincerely thank Asst. Prof. Dr. Onur Tümtürk and Assoc. Prof. Dr. Pelin Yoncacı Arslan for serving on my thesis jury. I truly appreciated the time and care they devoted to reading my work and the feedback they provided during the defense. Their comments and reflections were both encouraging and constructive, and I am grateful for their contribution to the development of this thesis.

I would also like to express my appreciation to Meral Özdengiz Başak, with whom I had the opportunity to work as a teaching assistant in the ARCH 401 and 402 design studios. It was a pleasure to take her studio as a student, and later to return as a teaching assistant. During this experience, I also had the pleasure of working with Ezgi İşbilen, Zeynep Ege Odabaşı, Melis

Sözen, and Burak Çelik in the fourth-year studio team. More than simply executing studio responsibilities, they demonstrated a strong commitment to fostering student growth.

I am especially thankful to Bilal Kabaş for his constant encouragement, mental support, and steady presence throughout this journey. He has been a source of inspiration during the writing of this thesis. I am deeply grateful for his presence by my side every step of the way.

Finally, I would like to express my deepest gratitude to my family, who have always stood by me with love and support. And to my beloved cats, Sushi and Şila, for providing comfort, joy, and quiet companionship during the long nights of writing and reflection. Thank you for always being by my side, sitting next to my laptop, throughout both my undergraduate and graduate years.

TABLE OF CONTENTS

ABSTRACT iii

ÖZET iv

ACKNOWLEDGEMENTS v

LIST OF FIGURES ix

LIST OF TABLES xiv

LIST OF ABBREVIATIONS xv

1 Introduction 1

1.1 Research Problem Statement 3

1.2 Aim and Scope of the Thesis 6

1.3 Methodology of the Research 7

1.4 Structure of the thesis 10

2 CARTOGRAPHIES OF URBAN BETWEEN CRITIQUE, REPRESENTATION AND TECHNOLOGY 11

2.1 Critical Cartography 11

2.1.1 Mapping as an Instrument of Power, Control and Space 14

2.1.2 Mapping as Resistance 18

2.1.3 Map Hacking 21

2.2 Representational Paradigms: Cartography, GIS and Beyond 24

2.2.1 Cartographic Semiosis: Maps as Sign System 29

2.2.2 Computational Epistemologies of the City: From Observation to Operation 32

2.2.3 Machine Semiosis: Algorithmic Reading and Visual Parsing 37

2.3 Street View Imagery as a Tool for Urban Operation 41

2.3.1 The Myth of Neutrality in Street View Imagery 45
2.3.2 Implementations of Street View Imagery 52

3 Karaköy Through Time: Empire, Republic and Redevelopment 60

3.1 Karaköy as a Site of Transformation 60
3.2 Late Ottoman Period 61
3.3 Early Republican 69
3.4 2000s and Beyond 72

4 Data Acquisitions, Inferences and Results 80

4.1 Framework of Visual-Linguistic Interpretation 80
4.2 Layering with Signification 91
4.2.1 Elemental Layer 93
4.2.2 Systemic Layer 95
4.2.3 Synthetic Layer 97
4.2.4 Presentational Layer 99
4.3 Results and Discussion 100
4.3.1 Implementation Outcomes from the Karaköy 100
4.3.2 Evaluation Metrics and Model Behavior 124

5 Conclusion 129

5.1 General Evaluation 129
5.2 Limitations of the method 131
5.3 Future Directions and Applications 134

REFERENCES 137

APPENDIX 145

LIST OF FIGURES

1.1 Methodology represented in three steps 9

2.1 The map shows the abstract lines which demonstrate the power aspect of space, North Carolina (Wood, 1992). 15

2.2 The figure illustrates the rational, geometric subdivision of land into uniform parcels, highlighting the visual logic of cadastral mapping as a tool of spatial governance. Adapted from Hildegard Binder Johnson (as cited in Pickles, 2004). 17

2.3 Anti-Eviction Mapping Project displacement mural (Kollektiv Orangotango, 2018). 19

2.4 596 Acres signage and community garden reclamation (Kollektiv Orangotango, 2018). 21

2.5 Anonymization in mapping platforms. Left: blurred faces in Karaköy. Right: obscured license plate. These show how platforms address privacy. 27

2.6 This figure illustrates that the cartographic symbol for a lake derives its meaning not from individual elements like lines or fills alone, but from their combined and context-dependent relationship as a culturally learned sign construct (Wood, 1992) 31

2.7 In Barthes's view, the signifier and signified form a sign, which myth then uses as a new signifier in a second-order system. He notes that describing this process as spatial is only a metaphor (Wood, 1992), redrawn from (Barthes, 2015)..... 31

2.8 Shannon's 1948 schematic of communication system, redrawn by Crampton. 34

2.9 Heatmap comparison between two urban scenes as interpreted by DINOv2. In the first row, the model clearly attends to familiar architectural features such as window bars and facade symmetry. In the second row, although there is a visible opening similar to a window, the model's attention is less focused and more dispersed..... 40

2.10 A Google Maps interface displaying blue-highlighted roads, indicating Street View availability, with Pegman ready to be placed. 41
2.11 The Street View interface as accessed through Google Maps..... 42
2.12 A Google Street View car equipped with a panoramic camera system used to capture 360-degree imagery for Google Maps (Google, n.d.)..... 43
2.13 Examples of specialized Street View capture devices adapted for diverse terrains, including a tricycle (2008), snowmobile (2010), and Trekker backpack (2012) (WIRED, 2022)..... 44
2.14 Alternative mobile platforms used in Street View projects, showing a camel-mounted camera in the Liwa Desert and a sheep-mounted system in the Faroe Islands (Parkinson, 2014; Vega, 2016). 44
2.15 The number of GSV panoramas available for New York City, shown by year (Curtis et al., 2013). 50
2.16 A spatio-temporal bus trip map to temporal flattening 51
2.17 Model attention visualized via CAMs (Raghu et al., 2023). 54
2.18 Detecting window-to-wall ratios for urban-level analysis..... 55
2.19 MIT Place Pulse dataset and the scores given in six dimensions (F. Zhang et al., 2018)..... 55
2.20 Urban renewal probability across the street network predicted by the GNN model. Road segments are color-coded by renewal intensity: green (¡0.4), yellow (0.4–0.6), orange (0.6–0.8), red (¿0.8). Data source: OpenStreetMap (Y. Zhang et al., 2023)..... 57

3.1 Miniature depiction of Istanbul from Matrakçı Nasûh's Beyân-ı Menâzil-i Sefer-i Irâkeyn (Dağdelen, 2021)..... 64
3.2 Goad's 1905 fire map shows building materials and functions west of Galata Bridge, near port structures (No:26) (Goad, 1905). 66
3.3 Timber Vitalis Han ve Ottoman Bank Building on Voyvoda Street (SALT, 2021). 67
3.4 An illustration of Galata's transformation as envisioned by Prost, taken from his 1/2000 scale Galata-Pera Plan. 70

3.5 A 1958 image capturing a demolition activity near Karaköy Square, extending toward Bankalar Street. Noradunkyan Han was demolished. (Eldem, 2000) 72
3.6 Satellite image of the Galataport project area along the Karaköy coastline, outlined in red (Şehir Planlama Müdürlüğü, n.d.). 75
3.7 The former Karaköy Passenger Terminal before its demolition as part of the Galataport redevelopment, representing Istanbul's mid-20th century port infrastructure (Şehir Planlama Müdürlüğü, n.d.). 76
3.8 The design proposal for the Peninsula Hotel, developed by the Hong Kong-based hospitality group (Şehir Planlama Müdürlüğü, n.d.). 77

4.1 Juxtaposing the Google Maps, OSM Data, and CAD file from Municipality in QGIS. 83
4.2 Occlusion check: intersection of line-of-sight from GSV point to building centroid 83
4.3 GSV points and occlusion-based exclusions based on building intersections 84
4.4 Categorization of data quality based on visibility conditions 84
4.5 GSV points every 15 meters around Galataport 86
4.6 GSV points every 15 meters on shoreline 86
4.7 Axis assigned to each point to match GSV images 87
4.8 Individual street views grouped into a temporal grid as ping. 88
4.9 Framework of Elemental Layer 94
4.10 Visual encoding of transformation types within the Elemental layer using triangular symbols. 95
4.11 Visual interpretation of transformation events through diachronic mapping of pings, enabling human perception of spatial proximities and emergent systemic patterns. 96
4.12 Unsupervised clustering of transformation data using HDBSCAN, producing data-driven groupings of events visualized as transformation clusters on the urban map. 96
4.13 Schematic of the Synthetic layer, where spatial clusters are linked with contextual data to support interpretive analysis of urban transformation. 98

4.14 Presentational Layer showing the cumulative GSV coverage of all years over a NolliMap basemap, with overlaid Boolnaz Signals and systemic inferences from clustering analyses. 100
4.15 Annual distribution of Google Street View images used in the study, showing variation in availability over time. 101
4.16 Distribution of Detected Transformations and Available Images per Year 102
4.17 Annual distribution of transformation types detected by the AI model between 2014 and 2024. 102
4.18 Yearly GSV Coverage and Transformation Types in 2014. Green indicates available GSV data, red indicates missing data, and orange triangles represent no-building areas. 104
4.19 Yearly GSV Coverage and Transformation Types in 2015. In the transformation map, all markers are shown as triangles, with their colors representing different types of change: blue indicates functional transformation, green indicates façade transformation, red indicates structural transformation, orange represents building removal, and purple denotes cases with insufficient data. 105
4.20 JW Marriott Istanbul Bosphorus construction phases on Kemankeş Street, one of the structure pings in 2015. 106
4.21 Yearly GSV Coverage and Transformation Types in 2017. 106
4.22 One of the depiction facade and function pings in 2017, 2018, 2019, and 2024 in Mumhane Street. 107
4.23 Yearly GSV Coverage and Transformation Types in 2018. 108
4.24 Galataport construction phases, captured across nine different years. 109
4.25 Yearly GSV Coverage and Transformation Types in 2019. 110
4.26 Hotel construction phases are completed along Hoca Tahsin Street, with corresponding structure pings detected in 2018 and 2019. 111
4.27 Yearly GSV Coverage and Transformation Types in 2020. 112
4.28 Yearly GSV Coverage and Transformation Types in 2021. 113
4.29 Yearly GSV Coverage and Transformation Types in 2022. 114

4.30 Yearly GSV Coverage and Transformation Types in 2024. 115
4.31 Facade transformation of a building between 2017 and 2024, with a major visible change occurring in 2024 through a pink exterior renovation. 115
4.32 Galataport shoreline as seen from the sea, comparing views from 2014 and 2024. 116
4.33 Clusters and noise points (outliers) of detected urban transformations in the Galataport area using HDBSCAN 118
4.34 Historical Value Distribution in the Study Area (Okyay & Binan, 2017). 120
4.35 Confusion Matrix of Predicted and Ground Truth 126
4.36 Spatial Distribution of Model Prediction Outcomes in Karaköy. 127

A.1 Composite visual layout of archival and GSV AND Mapillary images associated with multiple urban characters across different time periods and locations. 146
A.2 OSM Map shows the location of archival images. 147
A.3 Spatial distribution of transformation clusters produced by the AI-based HDBSCAN algorithm 147
A.4 JSON schema used to guide visual temporal reasoning and transformation classification. The prompt defines structured keys for image descriptions, temporal synthesis, and multi-label transformation output. 148
A.5 Condensed tree visualization of the AI-based HDBSCAN clustering output 149
A.6 Condensed tree visualization of the human-verified HDBSCAN clustering output. 149
A.7 A comparison of GSV and Mapillary imagery demonstrates how user-generated content can reveal an intermediate facade state that is not captured by GSV. This highlights the potential of platforms like Mapillary in counter-cartographic practices, where unofficial or overlooked spatial conditions are made visible through alternative forms of documentation. 150

LIST OF TABLES

2.1 Summary statistics for the first 10 selected U.S. cities (Kim & Jang, 2023) 49

2.2 Overview of urban transformation detection studies using street-level imagery. .. 58

4.1 Structure and Interpretation Rules of JSON Prompt 91

4.2 Accuracy of transformation Classifications 124

A.1 Detected Years of Transformation in Outlier Samples 145

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CAD	Computer-Aided Design
CAM	Class Activation Map
CNN	Convolutional Neural Network
DINOv2	Self-supervised Vision Transformer variant “DINO v2”
DPSM	Deep Pyramid Similarity Model
ETH	Eidgenössische Technische Hochschule Zürich (ETH Zurich)
GAN	Generative Adversarial Network
GIS	Geographic Information System
GPS	Global Positioning System
GSV	Google Street View
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
IoT	Internet of Things
JSON	JavaScript Object Notation
LLM	Large Language Model
MIT	Massachusetts Institute of Technology
MSVM	Markovian Support Vector Machine
NUS	National University of Singapore
ODbL	Open Database License
OSM	OpenStreetMap
PPUR	Patterns and Processes of Urban Redevelopment
QGIS	Quantum GIS
RGB	Red–Green–Blue colour model
ViT	Vision Transformer

VLM Vision-Language Model

XML Extensible Markup Language

YOLOv5 “You Only Look Once” object-detection model (version 5)

Chapter 1
Introduction

The increase in global population and the expansion of urban settlements have intensified the need for systematic and scalable tools to observe physical changes in the built environment. As urban areas grow in size and complexity, conventional techniques based on manual observation and field surveys face significant limitations in capturing the spatial and temporal dimensions of transformation. In response to these challenges, the development of urban data infrastructures and recent advancements in artificial intelligence have introduced new methods for documenting and analyzing urban change. Among these, street-level imagery platforms—particularly Google Street View (GSV)—have become widely used due to their extensive spatial coverage and accessibility. Originally designed for navigational purposes, GSV now functions as a visual database of urban form, offering high-resolution imagery from a human-eye perspective (Biljecki & Ito, 2021; Sun et al., 2022). Compared to aerial or satellite imagery, GSV provides more detailed information at the facade level and across multiple time points, enabling the detection of small-scale modifications such as alterations to storefronts, material surfaces, or building functions that are often not visible in traditional remote sensing data (Curtis et al., 2013; Y. Liu et al., 2025; Vallebueno & Lee, 2023).

Despite the spatial resolution and geographic coverage of GSV imagery, current computational approaches for analyzing urban transformation are predominantly designed to detect isolated visual elements at a single point in time. Commonly applied models—such as convolutional neural networks (CNNs) and vision transformers (ViTs)—are often used to extract visual indices including greenery, walkability, or perceived physical disorder (F. Zhang et al., 2018; Zhou et al., 2019). However, these methods typically process each image independently

and do not incorporate diachronic comparisons or scene-level continuity. As a result, changes that accumulate incrementally, such as alterations to signage, facade repainting, or functional repurposing of buildings, may not be systematically identified.

In addition to these methodological limitations, spatial and temporal inconsistencies in GSV image availability create further constraints for longitudinal analysis. The update frequency of GSV imagery is not uniform across geographic areas; socio-economic variables appear to influence the regularity with which certain neighborhoods are documented (Kim & Jang, 2023; Quinn & Alvarez León, 2019). Consequently, the resulting dataset may exhibit coverage imbalances that affect the generalizability and representativeness of analytical outcomes. These patterns reflect platform-specific update protocols and data acquisition strategies that shape the availability and visibility of urban imagery (Crampton, 2011). Under such conditions, GSV functions not only as a data source but also as a representational system with inherent structural characteristics that influence both what is recorded and how it can be interpreted.

Despite its growing potential, the integration of Google Street View and computer vision into urban research continues to face fundamental limitations. The scale and complexity of visual data pose significant challenges for interpretation, particularly in dense urban environments where transformation occurs rapidly and unevenly. Existing computational approaches often fall short in capturing temporally layered or scene-level changes, as they are typically designed for object-level detection without considering historical depth or spatial context. Furthermore, the epistemological foundations of these tools remain underexamined: both the visual datasets and the algorithms applied to them carry embedded assumptions and ontological constraints that shape what is rendered visible or legible. These issues highlight the need for analytical frameworks that are not only technically robust but also critically informed by an understanding of urban temporality, representation, and socio-technical mediation.

Responding to these gaps, this study proposes a computational and interpretive framework that combines visual language models (VLMs) with a domain-specific typology of transformation—structure, facade, and function. The proposed method integrates three interlinked layers: a computational layer that prepares temporally-aligned panoramas; an AI implementation layer

that parses visual inputs using structured JSON-based prompting; and an interpretive layer that situates detected transformations within socio-spatial contexts. The result is a multi-modal and multi-scalar method capable of capturing, classifying, and contextualizing micro-scale visual changes across time.

The empirical setting of this thesis is Karaköy, a historically layered district in central Istanbul characterized by overlapping regimes of infrastructure, commerce, and redevelopment. As a site that has witnessed intense and layered spatial trnsformations — from Genoese fortifications and Ottoman port functions to neoliberal redevelopment projects such as Galataport — Karaköy provides a compelling case for diachronic analysis (Eldem, 2000; Şehir Planlama Müdürlüğü, n.d.). Here, the proposed framework is operationalized to extract visual "pings" (discrete changes in image pairs) and to aggregate them into transformation "signals" that reflect broader urban dynamics.

By reading GSV imagery not as static evidence but as mediated representations conditioned by socio-technical infrastructures, this study aligns itself with critical cartography traditions (Harley, 1989; Wood, 1992) and computational urbanism (Crampton & Krygier, 2005). It examines not only what changes are visible, but how visibility itself is constructed—through platform design, algorithmic abstraction, and representational choices. Operationally, this thesis implements a three-layered pipeline that retrieves and processes street-view imagery, applies vision-language models to generate visual descriptions and classify transformation types, and spatially contextualizes the results to analyze urban change dynamics.

1.1 Research Problem Statement

Rapid urbanisation and the proliferation of digital imagery have made remotely sensed data indispensable for city research. While satellites and airborne sensors have historically been used to monitor land cover change, infrastructural development, and environmental patterns (Crampton, 2011; Navalgund et al., 2007), recent advances in digital mapping platforms have shifted attention to street-level imagery. Ground-based systems such as Google Street View

(GSV) provide billions of panoramas across more than 85 countries, offering highly detailed representations of the built environment at eye level (Biljecki & Ito, 2021; Sun et al., 2022). This archive enables planners and researchers to detect micro-scale transformations, evaluate policy interventions, and document the evolving dynamics of everyday urban life (Bahçekapılı, 2020; Curtis et al., 2013; Li, 2021).

In megacities, the scale and complexity of urban transformation make systematic, manual observation highly impractical. Tracking changes across thousands of facades, street segments, and building functions over time exceeds the capacity of individual observers and conventional fieldwork methods. The fragmented temporal coverage and irregular spatial dynamics of such environments require scalable and automated approaches. In this context, computer vision and remote sensing platforms—particularly tools like Google Street View—serve not as supplementary aids but as essential infrastructures for detecting, documenting, and analyzing spatial change at scale. Without such systems, many forms of transformation would remain beyond the reach of empirical analysis.

Most GSV research still relies on one-off snapshots processed with object-centric deep-learning models—typically convolutional neural networks (CNNs) or Vision Transformers (ViTs)—to score greenery, walkability, perceived safety, and other visual indices (Biljecki & Ito, 2021; F. Zhang et al., 2018; Zhou et al., 2019). A recent example by Vallebueno and Lee (2023) trains a YOLOv5 detector to flag potholes, graffiti, and broken windows, then aggregates these single-date detections into a street-level "urban-quality index" across three cities. While methodologically sophisticated, such pipelines treat each panorama as an isolated frame: they do not track whether a facade was repainted, a shopfront rebranded, or a vacant lot infilled over successive years.

Temporal reliability compounds the problem. Imagery dates can jump abruptly within the same road segment—sometimes every 25 m—introducing hidden discontinuities that corrupt longitudinal inference (Curtis et al., 2013). Nationwide data for 45 mid-sized U.S. cities show that 44% of home-to-work walking routes have no current coverage, and the mean age spread of images along a single path approaches seven years (Kim & Jang, 2023). These gaps are not

random: wealthier districts receive more frequent updates, while lower-income streets remain visually "stalled," perpetuating coverage bias (Quinn & Alvarez León, 2019).

Despite the growing use of computer vision techniques and street-level imagery in urban studies, three critical gaps continue to shape the limitations of current research: a scalability gap, a methodological gap, and an epistemological one. *(i) In large and complex urban environments, the spatial and temporal scale of transformation makes manual, human-led analysis infeasible.* Observing each change through fieldwork or individual annotation is not a viable strategy when thousands of building facades and streetscapes must be monitored across multiple time points. The inability to comprehensively observe and compare such large datasets without computational assistance creates a scalability bottleneck. Accordingly, the increasing reliance on automated, vision-based infrastructures such as Google Street View is not optional but necessary for enabling spatio-temporal urban analysis at meaningful resolution and extent. Despite the growing use of computer vision techniques and street-level imagery in urban studies, two critical gaps continue to shape the limitations of current research: a methodological gap and an epistemological one. *(ii) Methodologically, existing vision pipelines are largely confined to the detection of discrete objects in single-date panoramas and score-based comparisons*. These models, often based on convolutional neural networks or Vision Transformers, are designed to identify visual elements such as vehicles, trees, or signage. However, these methodologies do not extend to scene-level, temporally aligned detection. As a result, subtle yet meaningful transformations—such as changes in facades, the removal or replacement of signage, or shifts in building use—are rarely captured. Existing studies tend to focus on the extraction of visual indices rather than detecting or interpreting transformation events in relation to broader socio-economic processes. *(iii) Epistemologically, linking those visual changes to socio-spatial narratives often exposes a gap: the platforms and models that generate the data are themselves epistemic actors whose representational logics remain largely unexamined.* These three contributions are not equal in weight. The thesis primarily focuses on their intersection by modeling how AI-assisted change detection can function in complex and temporally layered urban environments. Google Street View imagery is produced through proprietary capture schedules, selective blurring, and uneven temporal–spatial coverage; machine-learning

models inherit those biases through their training sets and pre-defined classes. Consequently, any attempt to "ground" image-based detections in context must confront questions of data bias, missingness, and platform framing effects. In other words, the critique can no longer be limited to what a map (or an image) shows; it must also interrogate how these systems manufacture visibility in the first place. This study treats GSV and its attendant vision pipelines as representational infrastructures that are far from neutral data conduits, and argues that they require the same kind of deconstructive scrutiny long applied to traditional cartography, shifting the focus from representations to the underlying mechanisms through which meaning and visibility are constructed.

If the empirical and epistemological gaps outlined above remain unaddressed, significant yet incremental forms of urban transformation will continue to elude computational analysis. The absence of scene-level, temporally aware methods limits the capacity to trace change in contexts where formal records are lacking and transformation unfolds gradually. Moreover, reliance on unevenly updated imagery risks reproducing spatial biases and obscuring less legible urban conditions. This challenge is further compounded in megacities, where the rapid pace and density of urban change make detection increasingly difficult. Addressing these challenges requires both technical advances in temporal detection and a critical engagement with the representational assumptions of the platforms involved. This study proposes a framework that combines vision-language modeling with contextual interpretation to meet this dual imperative.

1.2 Aim and Scope of the Thesis

This study investigates how publicly accessible street-level imagery, with a focus on Google Street View, can be used to analyze urban transformation with Artificial Intelligence (AI). Instead of treating such imagery as a neutral or comprehensive record of the built environment, it is approached as a partial and contingent visual archive—one that is fragmented, selectively updated, and shaped by technical, institutional, and representational filters. The research aims to identify and track physical changes in urban form over time, while also examining how these changes can be computationally detected and what analytical and interpretive

limitations arise in the process. To this end, a methodological framework is proposed that combines AI-assisted change detection with critical reflection on the underlying assumptions embedded in both the image data and the algorithmic models. This thesis positions itself as a methodological intervention that asks what forms of spatial knowledge can be generated from street-level platforms such as GSV, what remains obscured or excluded, and how these insights might complicate or enrich conventional modes of urban analysis.

- **How can Google Street View be used not only as a visual archive of urban change, but also as a methodological tool for spatial investigation?**
- **What kinds of physical changes in the built environment can be detected through Street View imagery, and how can these micro-changes be aggregated to reveal broader transformation patterns?**
- **What are the potentials and limits of algorithmic processes in the interpretation of visual urban data?**

1.3 Methodology of the Research

This thesis develops *a multi-layered methodological framework* to detect and interpret facade-level urban transformations through Google Street View (GSV) imagery. Rather than treating these images as neutral representations, the approach considers them as socially and technically mediated documents that encode both visibility and omission.

To structure the interpretation of visual change, the study adopts a domain-specific ontology consisting of three transformation categories: facade, structure, and function. This typology guides both the machine-level parsing of multimodal outputs and the human-led interpretation of urban processes. Facade changes include alterations to materials, colors, signage, and surface-level aesthetic elements; structure refers to architectural modifications such as demolition, extension, or new construction; and function captures shifts in building use, ownership, or public/private designation. This tripartite classification enables the model to not only detect change but to articulate its implications within the built environment.

At its core, the methodology proceeds through three interrelated layers:

- **Computational layer:** As visualized in the first module of the Figure 1.1, this layer begins by acquiring the nearest GSV point for each building centroid. Historical panoramas are retrieved, aligned by timestamp and heading, and stitched into a temporal grid. This grid becomes the basis for detecting facade-level change across time. All images are renamed using a strict filename schema that encodes location and time metadata, creating a searchable and auditable archive of urban viewpoints.

- **AI implementation layer (within computational scope):** As shown in the second module (Figure 1.1), stitched multi-year image grids are converted into structured prompts in JavaScript Object Notation (JSON) format.[1] These prompts are fed into a multimodal language model, which returns a standardized JSON output containing visual descriptions and a predicted transformation type.[2] The classification is guided by a tripartite ontology—structure, facade, and function—which organizes the model's interpretation into semantically meaningful categories.

- **Interpretive layer:** Detected instances of visual change are interpreted in relation to their spatial proximity and morphological coherence within the surrounding urban fabric. Rather than treating each observation as an isolated event, the analysis considers how these transformations co-occur and interact within localized contexts. To further enrich this reading, the analysis draws on contextual and socioeconomic conditions specific to each locality. Such an approach enables visual transformations to be situated within the political, economic, and cultural forces that structure urban change. This approach aligns with Denis Wood's conceptualization of mapping as a meaning-making process.

[1] JSON (JavaScript Object Notation) is a lightweight, human- and machine-readable data format that organizes information into key–value pairs. In this study, it is used to structure model inputs and outputs, such as visual descriptions and transformation types, in a consistent and queryable form.

[2] A multimodal language model is an artificial intelligence system that can interpret and generate responses based on both visual and textual inputs.

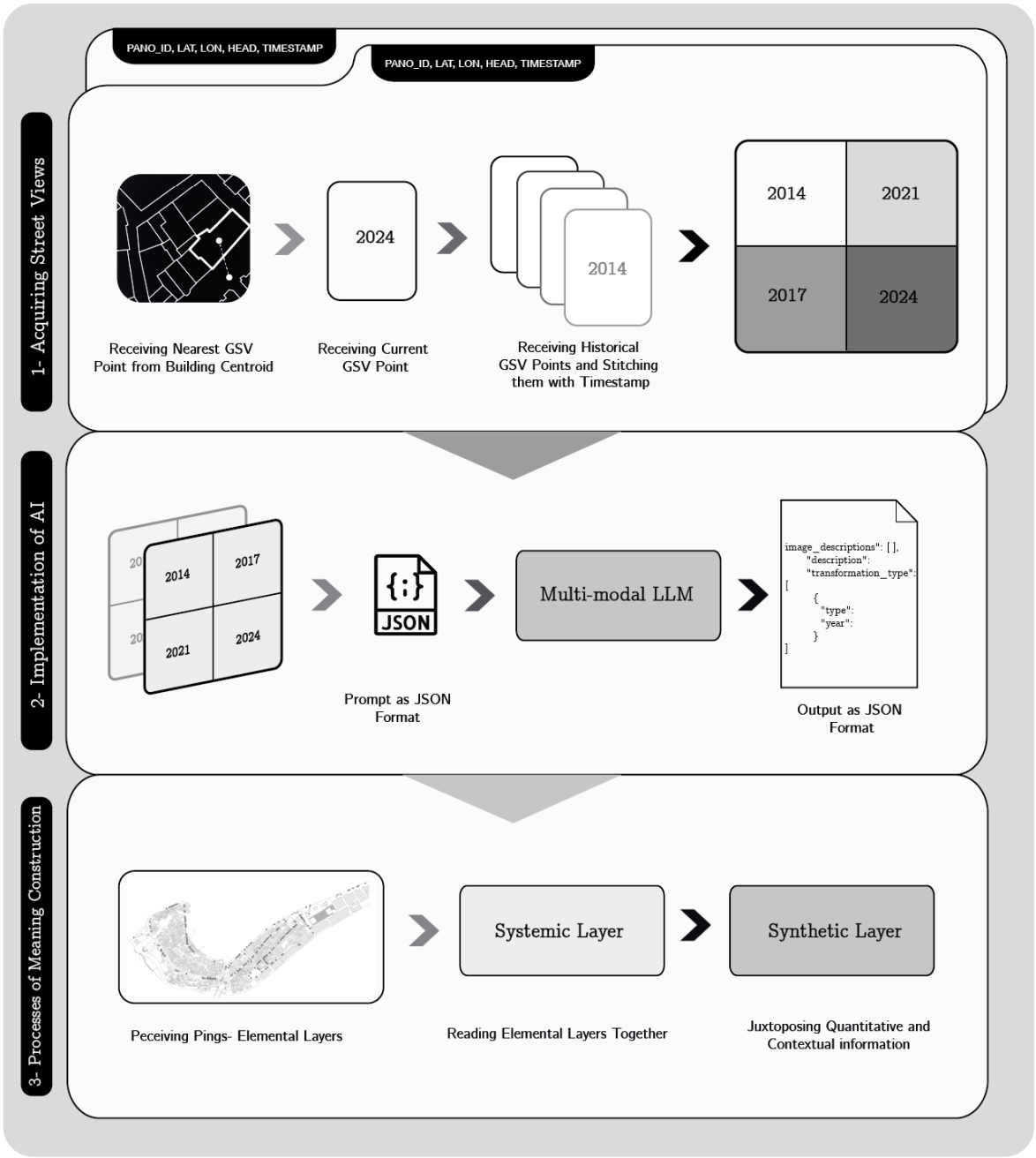


Figure 1.1: Methodology represented in three steps

1.4 Structure of the thesis

This thesis is organized into five main chapters, each contributing to the development of the theoretical foundation, empirical analysis, and methodological implementation of the research.

The second chapter reviews relevant literature in the fields of critical cartography and urban representation. It examines how mapping practices function not only as technical tools but also as frameworks that shape spatial perception. The discussion draws on key theoretical contributions by scholars such as Denis Wood, J.B. Harley, and Michel Foucault. The chapter also explores how digital platforms—particularly Geographic Information Systems (GIS), remote sensing, and computer vision—have altered the production and interpretation of spatial data, with a specific focus on Street View imagery.

The third chapter provides a historical overview of the case study area, Karaköy. It traces the spatial and functional evolution of the neighborhood across different historical periods, including its Byzantine, Genoese, Ottoman, Republican, and contemporary phases. The chapter synthesizes historical maps, planning documents, and urban redevelopment records to contextualize Karaköy as a site of continuous transformation and layered morphology.

The fourth chapter presents the methodological framework of the study. It describes the data acquisition process using Google Street View (GSV), including panorama alignment, stitching, and metadata structuring. The chapter introduces the computational pipeline based on a multimodal vision-language model and outlines the typology of transformation types—facade, structure, and function. Additionally, it explains the four-layered interpretive model—elemental, systemic, synthetic, and presentational—which is applied to analyze the spatial distribution and contextual significance of transformation events.

The fifth and final chapter discusses the results of the study and assesses the performance of the proposed method. It identifies the strengths and limitations of the approach, particularly in terms of model accuracy, spatial resolution, and representational scope. The chapter also proposes potential areas for future research, including methodological research.

Chapter 2

CARTOGRAPHIES OF URBAN BETWEEN CRITIQUE, REPRESENTATION AND TECHNOLOGY

2.1 Critical Cartography

Throughout history, maps serve functions far beyond navigation or record keeping. Not only do they mirror the physical world, but they also shape ways of imagining space, exercise power across territory, and legitimize or contest land claims through mapmaking. Prevailing ideologies are reinforced, spatial hierarchies are drawn, and stories useful to political, economic, or institutional interests are supported by these seemingly neutral documents (Crampton & Krygier, 2005). Behind every neat border and symbol, there are deliberate decisions about what information gets included, what details are excluded, and how different elements are visually related to one another. As Harley (1989) reminds readers, authors, subjects, and themes always guide maps, and silence matters as much as presence.

The authority of maps rests on their talent for presenting partial views as though no pieces are missing. Denis Wood calls this effect the *naturalization of the map*, because images shaped by cultural and institutional interests are received as plain facts (Wood, 1992). Through what he names *double-coding*, maps function as both visual aids and ideological tools. Distortion creeps in with every projection, scale, and simplification, even when clarity is desperately sought. As Monmonier notes, simplification is not judged a defect; instead, it is treated as a structural necessity that unavoidably shifts meaning (Monmonier, 2018). Lefebvre's notion of *representations of space* supports this critique, because spatial designs drafted by planners, architects, and bureaucrats gain legitimacy once they are mapped (Lefebvre, 2012). Although maps are not

Lefebvre's main subject, his ideas help explain how authority is lent to spatial arrangements. A similar point is emphasized by John Pickles, who shows that mapping practices validate dominant social relations and wider systems of control (Pickles, 2004).

Long-standing histories once traced a smooth path toward ever more scientific and accurate depictions. In recent decades, however, it has been demonstrated that meaning, representation, and authority have always been embedded in cartographic work. From sacred cosmographies to imperial surveys, maps have been used to manage resources and strengthen control (Crampton & Krygier, 2005; Harley, 1989; Pickles, 2004).

A decisive moment was reached during World War II, when Arthur Robinson was enlisted by the U.S. Office of Strategic Services. Tests on projections, symbol hierarchies, and color contrasts were carried out so that complex intelligence could be shown clearly. On that basis, the claim was made by Robinson that perceptual science, not artisanal intuition, should guide cartography. After the war, this stance was shared through The Look of Maps and Elements of Cartography, and a generation of mapmakers was trained to view their craft as technical and objective (Crampton & Krygier, 2005).

By the 1980s and 1990s, the idea that maps simply offered objective views of space was being challenged by more critical perspectives. Influenced by Foucault's theories on power and knowledge and Derrida's ideas about textual interpretation, J. B. Harley argued that maps should not be seen as neutral representations. Instead, he suggested they be read as cultural texts shaped by particular ideologies and historical contexts (Harley, 1989). Denis Wood extended the claim, insisting that maps do not merely reflect space but actively create spatial knowledge that serves institutions (Wood, 1992). A wider critique was developed by Jeremy Crampton and John Pickles, who linked mapping to colonialism, capitalism, and state formation. Borders have been naturalized, unequal land division has been upheld, and dominant spatial imaginaries have been reproduced through cartographic practice (Crampton & Krygier, 2005; Pickles, 2004).

Critique is not the search for mistakes; it is the effort to expose the assumptions that organise every system of knowledge (Crampton, 2011). The scholars cited in the previous paragraph pursue exactly this aim, demonstrating that maps rest on historically and culturally

specific conventions rather than neutral facts. From this perspective, cartography becomes an act of critique in its own right, because each decision about what to draw and how to draw it simultaneously questions and reinforces the authority that underwrites it. Recognising this dynamic reveals that knowledge is never simply objective; it takes shape and gains legitimacy only within particular social and temporal contexts. This is the foundation upon which critical cartography builds. Rather than taking the map for granted, it examines the deeper logics and embedded power relations that make mapping itself possible.

Critical cartography has shifted attention from technical mistakes on maps to the underlying assumptions that make mapping possible in the first place. Mapping is no longer seen as a neutral or purely technical task. Instead, it is understood as a political and epistemological practice that shapes how we perceive and organize space. Foucault's concept of power-knowledge shows how maps are closely tied to systems of classification, surveillance, and spatial control (Foucault, 2020). They do not simply describe space; they help produce it. Denis Wood captures this idea by describing the modern condition as a "map-immersed world," where maps are not just representations but shaping forces in everyday environments (Wood, 1992). In this context, maps influence what we see, how we move, and how we imagine change. They structure reality rather than merely reflect it.

In his later writings, Harley (1989) clarified that maps should not be viewed as objective representations, but rather as narratives shaped by institutional power, cultural norms, and shared understandings. Drawing on Foucault's discourse theory, he argued that cartography plays a key role in determining which spatial claims are accepted as truth. Derrida's deconstruction was used to reveal silences, omissions, and rhetorical choices in maps, but Harley favored Foucault's broader focus on the structural operations of power and knowledge. He proposed a three-part reading: first, maps function as discourse governed by implicit rules; second, their textual and persuasive features must be examined; and third, maps must be situated within formations of power-knowledge. This perspective challenges the idea of cartography as a product of uninterrupted scientific progress, presenting it instead as a professional myth that legitimizes its authority. Since the nineteenth century, scientific rationalism and bureaucratic procedures have marginalized artistic and narrative expressions in mapping. Ultimately, Harley

emphasized that trust in maps stems not from their accuracy, but from their alignment with dominant social orders (Harley, 1989).

Taken together, these critical perspectives suggest that maps should be understood not as passive depictions of spatial reality, but as practices that actively contribute to its construction. By examining the ideological, institutional, and discursive conditions under which mapping takes place, critical cartography highlights the political dimensions that shape spatial representation. Maps influence how territories are perceived, organized, and negotiated, and thereby intersect with broader systems of power. Recognizing these dynamics is important for reassessing how spatial knowledge is generated, interpreted, and contested in the present. Whenever meaning is mediated through technological, symbolic, or procedural systems, representations require close and critical interpretation. What is rendered visible or meaningful depends on the frameworks through which space is structured. In this light, critical cartography provides a conceptual approach for making current spatial arrangements more transparent and subject to examination.

2.1.1 Mapping as an Instrument of Power, Control and Space

Maps are often regarded as straightforward and reliable representations of space. Yet, they do not always reflect the full complexity of reality. Rather than presenting a definitive ground truth, maps are shaped by relations of power, systems of control, and intentional design. Even the internal components of mapping, such as signage, temporality, and the selective inclusion or exclusion of information, reflect subjective choices. Although mapping practices are often regarded as authoritative, they are ultimately representational systems that participate in the continuous construction and reproduction of space (Harley, 1989; Monmonier, 2018; Pickles, 2004; Wood, 1992).

According to (Lefebvre, 2012), such representations of space refer to the conceptualized spatial forms generated by planners and architects. These spaces are typically defined by abstraction order and control. While Lefebvre does not specifically address the role of cartographers, mapping can be understood as similarly contributing to the production of conceptualized

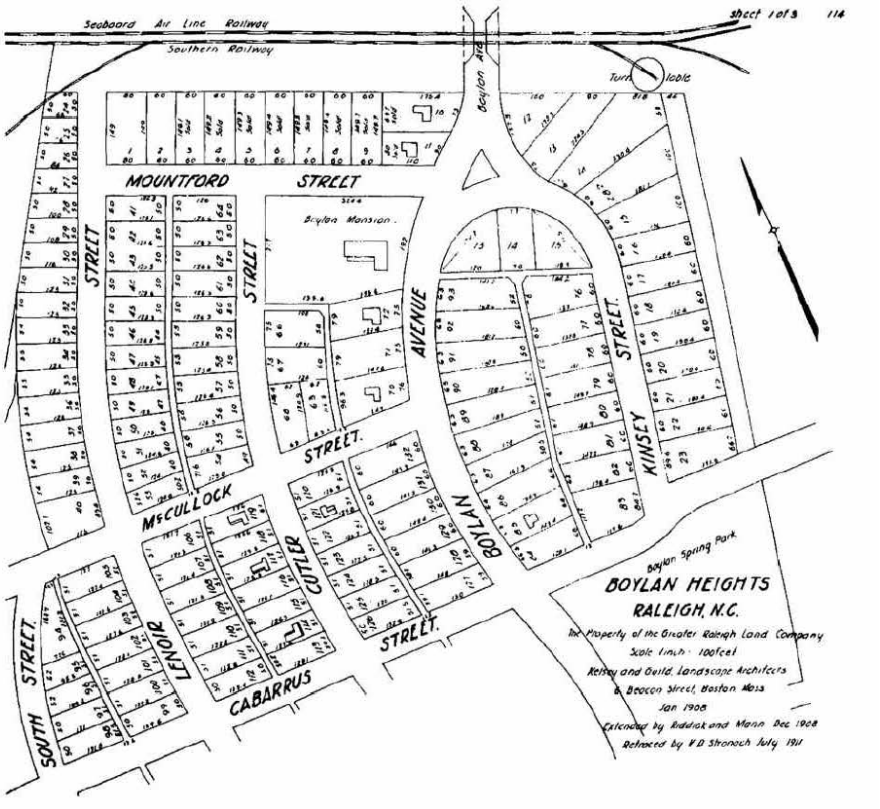


Figure 2.1: The map shows the abstract lines which demonstrate the power aspect of space, North Carolina (Wood, 1992).

space. It operates as a tool through which dominant ideologies are embedded within the built environment (Lefebvre, 2012).

An illustrative case is presented in Denis Wood's detailed analysis of a neighborhood in Raleigh, North Carolina. In this example, a residential area with privately maintained streets was incorporated within the official boundaries of the city. Although the municipality explicitly declined to provide services such as road maintenance, the act of mapping the neighborhood as part of the city effectively assigned it a new administrative identity. Despite the absence of direct state intervention or infrastructural integration, residents became subject to municipal taxation and other forms of local governance. What is especially significant is that this inclusion was not driven by any physical transformation on the ground, but was instead enacted through cartographic representation (Wood, 1992).

Wood (1992) uses this case to emphasize that maps do not merely reflect pre-existing spatial realities but instead operate as instruments that actively construct and enforce legal and

political frameworks. By delineating boundaries, maps make relationships of ownership, obligation, and authority both legible and enforceable. These outcomes are not derived from the inherent qualities of the land itself, but from the symbolic and administrative power embedded in the act of mapping. Administrative categories such as taxation, jurisdiction, and service provision emerge not from an absolute or neutral conception of space but from cartographic inscription that assigns them institutional meaning. In this way, spatial authority functions performatively: space becomes governable or taxable not because of its material characteristics, but because it is defined as such through representation. This logic is captured by the Latin phrase *quia ego nominor leo*—"because my name is lion"—which, as analyzed by Roland Barthes, illustrates how forms emptied of their historical content can assert authority through their own internal logic. Such phrases, once reframed as neutral or natural, carry legitimacy by virtue of their structure, not substance (Barthes, 2015). Similarly, as Denis Wood has argued, the map produces spatial realities that would not otherwise exist (Wood, 1992), shaping both abstract systems of governance and the lived experience of everyday environments.

A second example that illustrates the political function of mapping is the concept of *"gridded lands"*, as discussed by Hildegard Binder Johnson in Order upon the Land in Figure 2.2. This concept refers to the geometric division of land through cadastral mapping, especially during the territorial expansion of the United States in the eighteenth and nineteenth centuries. In this process, land was transformed from a lived, irregular, and often communal landscape into standardized parcels defined by straight lines and uniform measurements. The goal was to make land measurable, visible, and exchangeable. As a result, this transformation supported the needs of a growing capitalist economy and a centralizing state authority. Instead of responding to natural topography, historical land use, or cultural meaning, the new system imposed a rational and abstract logic onto space (Pickles, 2004).

This process replaced local spatial practices, such as oral boundaries, flexible land-sharing systems, and indigenous mapping traditions, with fixed and formalized units of property. These new units could be taxed, sold, inherited, or mortgaged. In this sense, mapping was not only about describing land, it was about producing land in legal and economic terms. The cadastral grid became a foundation for property law and market exchange. This new way of organizing

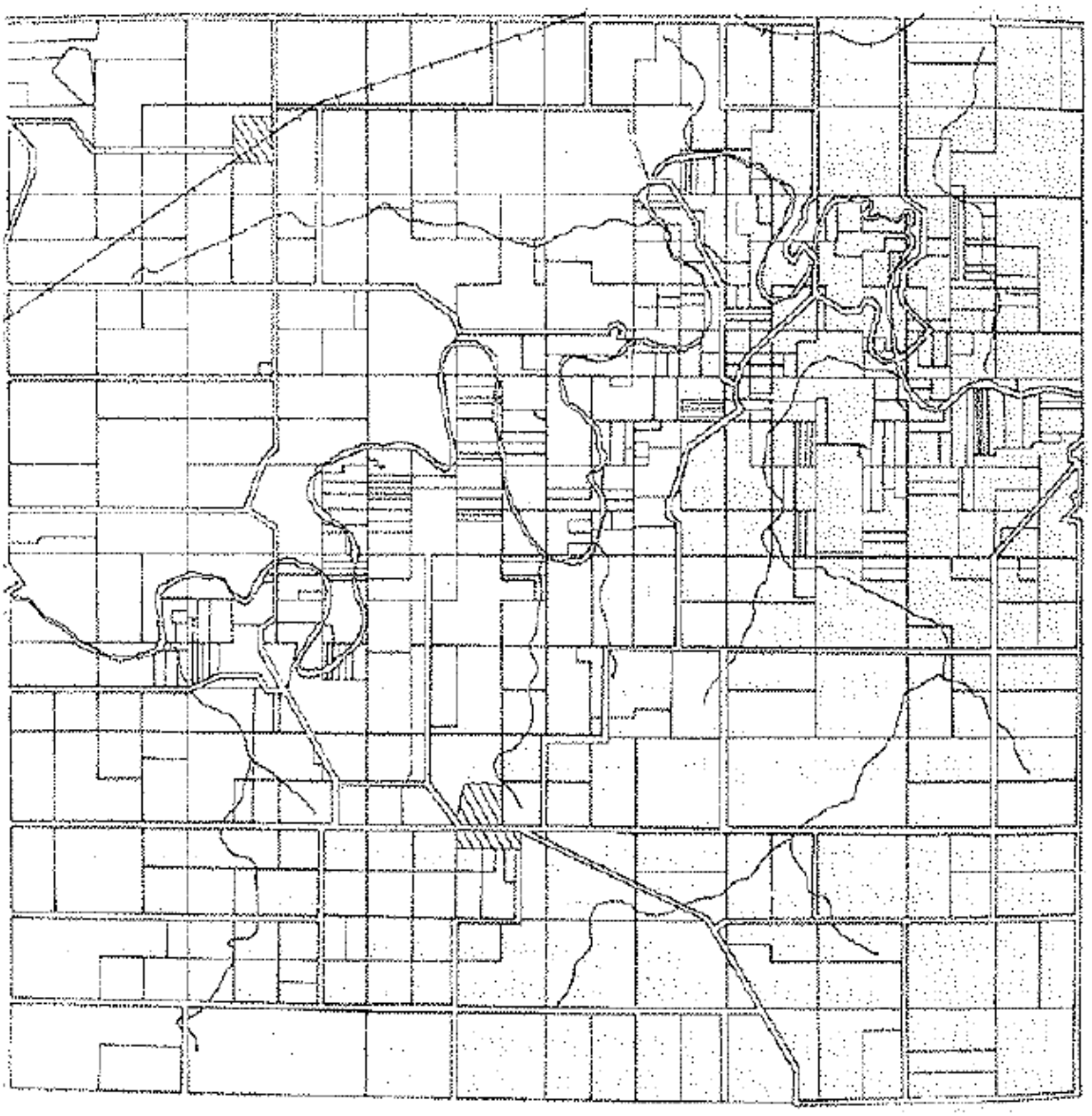


Figure 2.2: The figure illustrates the rational, geometric subdivision of land into uniform parcels, highlighting the visual logic of cadastral mapping as a tool of spatial governance. Adapted from Hildegard Binder Johnson (as cited in Pickles, 2004).

space allowed governments to manage land at a national scale, investors to calculate risk and profit, and settlers to claim ownership backed by the state. It also enabled the displacement of populations whose land use practices were incompatible with the new grid logic (Pickles, 2004).

This process replaced local spatial practices, such as oral boundaries, flexible land-sharing systems, and indigenous mapping traditions, with fixed and formalized units of property that could be taxed, sold, inherited, or mortgaged. Mapping played a central role in this shift by redefining land in legal and economic terms. The cadastral grid became a key mechanism, serving as a basis for property law and a tool for market exchange. The use of grids was not a neutral technical method; it functioned as a spatial strategy of governance. By imposing standardized geometries and drawing fixed property lines, the grid established a new legal and visual order that made land more ”legible” to state institutions. Land could now be seen, categorized, and acted upon through centralized systems of administration. In this way, maps did not simply depict space but participated in its transformation. What was included gained institutional meaning, while what was excluded risked being removed from official recognition altogether (Pickles, 2004; Wood, 1992).

When reading the map of a city—especially in contexts shaped by capitalist development—it is important to remember that neatly divided parcels reflect not only spatial order but also a shift in how space is conceptualized. Such representations point to an understanding of space that is removed from everyday practices and redefined within an abstract system where land becomes measurable, transferable, and subject to ownership. In this sense, what is shown on the map should be considered not just in terms of what it depicts, but in terms of the logic through which it organizes and enables certain kinds of spatial transformation.

2.1.2 Mapping as Resistance

Critical cartography builds on a simple yet significant idea: maps are often seen as representations of spatial power, but they can also participate in its production and maintenance. This perspective laid the groundwork for what Nancy Peluso later described as counter-cartography

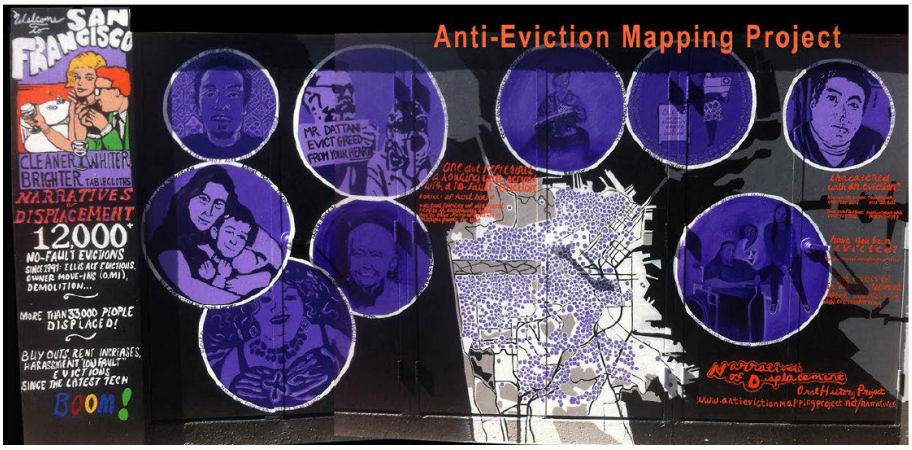


Figure 2.3: Anti-Eviction Mapping Project displacement mural (Kollektiv Orangotango, 2018).

(Kollektiv Orangotango, 2018). Years earlier, Harley (1989) and Wood (1992) had already argued that maps created by states and corporations work to make control over land, property, and people appear natural and unquestioned. Their critiques opened a space—both conceptual and political—for mapmaking to function as a form of resistance.

Counter-mapping, in this sense, goes beyond the idea of alternative cartography and becomes a form of resistance. It can be understood as the performative dimension of critical cartography, as it challenges dominant mapping practices through direct spatial engagement and by reclaiming the authority to represent. In this way, critique is not confined to commentary; it takes shape through action and representation. Counter-mapping becomes a practice of opposition, embedded in both method and intent (Kollektiv Orangotango, 2018).

The connection between critical cartographic thought and political practice becomes concrete in projects like the Anti-Eviction Mapping Project (AEMP) in Figure 2.3. AEMP began in San Francisco in response to increasing evictions linked to gentrification and urban redevelopment. Its work extends beyond standard geospatial tools by integrating eviction records, personal narratives, visual media, and collective learning processes.. Evictions are not treated as isolated events; they appear as outcomes of broader systems of exclusion. By layering demographic shifts, personal accounts, and acts of resistance, AEMP's maps trace how displacement follows patterns shaped by race, gender, and class. These maps are not neutral representations. They are built to confront, to record harm, and to retain what planning processes frequently ignore: the lived conditions of removal (Kollektiv Orangotango, 2018).

Where AEMP draws attention to forced removal, 596 Acres turns to what remains out of reach in Figure 2.4. It focuses on the public lands of New York City, parcels owned by the city but left unused and obscured by layers of administrative complexity. Through scraping municipal data and building interactive maps, 596 Acres makes these sites visible to the people living near them. Each entry on the map is more than a data point; it signals a possibility. Physical signs with QR codes are fixed to fences, shifting the meaning of these spaces. What looked like a forgotten lot becomes something legible, and open to use (Kollektiv Orangotango, 2018; Segal & Acres, n.d.).

The tools mobilized by 596 Acres to facilitate community-led land reclamation are modest in form yet transformative in effect: hand-posted signs, neighborhood meetings, and editable online maps. These grassroots instruments serve not only as communicative media but as interventions that reshape public perception. An empty plot is no longer seen as wasted space but as one full of potential, ready to become a garden, a place to gather, or a space for learning. Legal procedures matter here, not just images. The map becomes part of a larger socio-legal process that includes permits, deliberations, and collective decision-making. In this sense, 596 Acres treats mapping not merely as representation but as activation—a participatory strategy for reclaiming and reimagining public space (Kollektiv Orangotango, 2018; Segal & Acres, n.d.).

In this context, counter-maps operate on at least two interconnected levels. On a practical level, they serve as tools that support legal proceedings, inform policy debates, and mobilize public campaigns. On a conceptual level, they cultivate a reflexive mapping consciousness—an awareness, as Pickles argues, that maps do not merely represent the world but actively participate in its construction (Pickles, 2004). From this perspective, mapping workshops are not simply spaces for acquiring technical skills; they become political arenas in which claims to land are articulated, demands for visibility are asserted, and collective forms of action are envisioned. At the same time, many counter-mapping initiatives are acutely aware of the risk of generating new forms of exclusion, even as they seek to challenge existing ones. In line with Harley's reminder that every map inevitably leaves something out, these projects often adopt principles of open access, adaptability, and shared authorship as strategies for accountability



Figure 2.4: 596 Acres signage and community garden reclamation (Kollektiv Orangotango, 2018).

and inclusivity (Harley, 2018). Thus, mapping is redefined not merely as an act of representation but as a form of intervention, negotiation, and social transformation. These examples demonstrate that mapping operates as a form of spatial and political mediation, shaping access, visibility, and claims to space when enacted through tools such as signage, digital interfaces, or participatory practices.

2.1.3 Map Hacking

In the early 2000s, the practice of cartography began to change as new digital technologies became accessible to ordinary people. Until then, making maps required professional knowledge, costly software, and authoritative access to government data. That changed when companies like Google opened their mapping platforms through Application Programming Interfaces (APIs), allowing users without formal training to build interactive maps using little more than a browser and basic code implementation. The rise of *"Maps 2.0"* marked a turning point, during which mapping was increasingly taken up by amateur users and redirected from state or corporate usage toward more personal and local concerns. [1] This decentralization of cartographic

[1] Maps 2.0 describes a new way of making maps where non-experts can create and edit them online. It became possible through open tools like Google Maps API and platforms that support user participation.

authority coincided with the wider development of the *geoweb*, a collection of web-based services and platforms explicitly built around geographic information, facilitating the widespread distribution of spatial data and enabling collaborative and participatory map-making practices such as crowdsourcing (Crampton, 2011).[2]

As mapping tools and geographic data became more accessible, the discourse surrounding map-making began to shift and was increasingly shaped by the language of computing. The term *hacking*, which is often associated with unauthorized access to computer systems, was instead reclaimed in reference to its original meaning. In its earlier usage, hacking described a practical and exploratory approach that involved taking systems apart to understand how they worked and experimenting with their components in unfamiliar or creative ways (Crampton & Krygier, 2005). In this context, *map hacking* came to describe the inventive reuse of digital mapping tools. Instead of building maps entirely from the ground up, map hackers combined existing platforms and datasets, often crowdsourced from multiple contributors, to produce something new and contextually meaningful. (Crampton & Krygier, 2005) viewed this method as a low-cost, accessible approach that allowed individuals with basic coding skills to develop spatial tools with immediate civic applications.

McConchie (2015) expanded this conversation by introducing the notion of *hacker cartography*, a term capturing the cultural and political implications of these emerging practices. Hacker cartographers generally work outside traditional hierarchies, blending technical skill with collaborative, community-driven energy. OpenStreetMap (OSM) exemplifies this approach. Created in 2004 in reaction to restrictive government and corporate data policies, OpenStreetView (OSM) invites volunteers worldwide to collect and share spatial data using aerial imagery tracing or local field notes. This crowdsourced data is published under open licenses such as Creative Commons or Open Database License (ODbL), which ensures free reuse (Crampton & Krygier, 2005). McConchie (2015) emphasizes that OSM is more than just a technical project; it constitutes a political statement supporting openness, shared labor, and bottom-up knowledge production within the broader context of the Geospatial Web ”GeoWeb”.

[2]The GeoWeb is the use of the internet to access, create, and share geographic information. It connects maps with online platforms and allows users to interact with spatial data.

Early examples of map hacking demonstrated this potential. In 2005, Paul Rademacher created HousingMaps by linking Craigslist real estate listings with Google Maps, producing a geographically searchable database of apartment listings. Around the same period, Adrian Holovaty launched Chicago Crime, a platform mapping police arrest data across city neighborhoods (Crampton, 2011). Both projects used publicly available data and existing digital mapping tools to produce innovative visualizations previously unavailable through traditional workflows. These examples did not represent sabotage or illegal activity; rather, they illustrated the transformative potential of map hacking to convert passive data into dynamic, interactive resources closely aligned with local interests and public needs. Furthermore, they reflected the main principles of crowdsourcing by aggregating and visualizing information contributed by multiple sources and making it accessible through the GeoWeb. These developments were made possible by the growing availability of Application Programming Interfaces (APIs), customizable maps, and interactive layers. These tools allowed users to combine and reinterpret data in new ways, beyond its original context. Technologies such as Extensible Markup Language (XML) and API protocols enabled different software systems to communicate with one another, creating a foundation for more open and flexible forms of digital mapping. One key outcome of this infrastructure was the emergence of map mashups, which are web-based visualizations that integrate multiple datasets into a single interactive map (Crampton & Krygier, 2005; McConchie, 2015).

However, the ideals and practices of map hacking are not without tension. McConchie (2015) notes that while hacker cartography strongly emphasizes open participation, it also raises significant questions regarding data accuracy, reliability, editorial consensus, and long-term maintenance. At the same time, major corporations initially distant from these grassroots experiments began incorporating similar methods, appropriating volunteer labor and data contributions. Platforms such as Google increasingly draw on user-generated content and crowdsourced data, but often without offering reciprocal openness or unrestricted access in return. This shift poses risks related to the enclosure of publicly contributed data, where initially open resources become enclosed again under proprietary ownership. According to Crampton and Krygier (2005) and McConchie (2015), the central questions are political rather than purely

technical. Map hacking and hacker cartography compel us to critically examine who controls geographic information, whose experiences and perspectives are made visible, and under what conditions spatial data is produced, shared, and circulated within the evolving geoweb.

2.2 Representational Paradigms: Cartography, GIS and Beyond

The representation of space is not governed solely by technical instruments; rather, it often operates as part of broader strategies involving surveillance, regulation, and control. Since the French Revolution, modern techniques of governance have established various regimes of visibility to count, classify, and define individuals. Tools such as population censuses, identity cards, fingerprint systems, and anthropometric measurements enable not only the collection of biological data but also the spatial tracking and regulation of bodies. This process of rendering individuals visible intensifies during the First and Second World Wars and connects to the emergence of modern capitalism and productivity (Crampton, 2011). As a result, space is no longer simply inhabited. It is mapped and processed as a data object.

Within the broader shift in cartographic logic, cartographic practices are also transformed. Initially employed as tools of observation, maps gradually reposition themselves at the center of operational and algorithmic processes. Particularly in the second half of the twentieth century, influenced by the development of information theory and digital computing, mapping is no longer understood as just a representational act. Through the works of figures such as Boole (1847), Leibnitz (1703), and Shannon (1948).[3] Maps are reconceived as devices capable of classification and intervention. Such a reconceptualization leads to a change where observation is replaced by operation as the main cartographic paradigm.

The shift toward operational cartography is most apparent with the rise of Geographic Information Systems (GIS). By integrating Internet of Things (IoT) devices, Global Positioning System (GPS) technologies, and remote sensing tools, spatial data are now processed instantly

[3]Leibniz developed binary logic as a universal language of reasoning; Boole formalized it into a symbolic system of logic; and Shannon operationalized it within electronic communication systems—together laying the intellectual groundwork for the transformation of spatial representation into a computable form.

and stored in digital systems.[4] The urban environment is reimagined as a flexible and searchable database. However, the emphasis on accuracy and precision often hides the assumptions built into technological infrastructures such as GIS platforms and spatial data frameworks.. Maps begin to function as tools that define what counts as a road, a residence, or a potential threat. As Crampton (2011) notes, the use of cartographic systems increasingly connects to practices of securitization.

The role of visual technologies in this process becomes increasingly pronounced. With the integration of computer vision systems, the act of mapping is delegated to machine-based perception.[5] Through the use of machine learning techniques, raw visual data are transformed into spatially meaningful categories. Systems such as facial recognition, object detection, and spatial annotation are employed to extract and classify features from imagery.[6] In doing so, these systems contribute to the reinforcement of spatial boundaries and culturally embedded assumptions. Decisions regarding what becomes visible, what is categorized, and what is excluded from representation are shaped by algorithmic models that are trained on selective, incomplete, or biased datasets.

Through these advancements, the relevance of Joe Painter's concept of *cartographic anxiety* becomes more pronounced (Crampton, 2011). The map is no longer perceived as a neutral window onto the world. Instead, it is understood as playing a constitutive role in the production of space. Decisions about who appears where, how space is classified, and what is rendered absent are produced through cartographic techniques infused with technical, political, and cultural assumptions. As a result, the map functions simultaneously as a tool for generating knowledge and as a spatial medium through which power relations are reproduced.

[4]Internet of Things (IoT) devices are physical objects equipped with sensors, software, and connectivity that enable them to collect and share data through communication networks. GPS (Global Positioning System) technologies use satellite signals to determine accurate geographic locations on Earth. Remote sensing tools are instruments placed on platforms such as satellites, aircraft, or drones, which gather information about the Earth's surface without physical contact, typically using electromagnetic or thermal data.

[5]Computer vision systems refer to computational techniques that allow machines to extract, analyze, and interpret information from visual data such as images or videos. These systems aim to replicate or approximate human visual understanding for tasks ranging from simple recognition to complex spatial reasoning.

[6]Facial recognition identifies individual faces from images or videos; object detection locates and classifies multiple items within an image; spatial annotation refers to labeling parts of an image with spatial metadata, often for training or analysis purposes.

Cartographic anxiety, as defined in critical cartography, refers to two related concerns. The first is the concern that mapping may unintentionally reinforce systems of power and control. The second is the epistemological uncertainty that arises from understanding maps not as neutral tools, but as representations shaped by cultural and political frameworks (Crampton, 2011). In the context of artificial intelligence and widespread surveillance, the anxiety takes on new dimensions. Mapping is no longer carried out just by human intelligence. Algorithms now play a decisive role in determining what is made visible, how it is categorized, and what becomes actionable, often through models trained on incomplete, selective, or biased datasets. Visibility is frequently assigned in advance, and individuals or groups may be labeled as risky or marginal without their awareness or consent.

What further complicates the relationship between spatial data collection and individual privacy is that platforms like Google Street View and Mapillary record ordinary urban environments where people's homes, vehicles, and daily routines become part of massive visual datasets. These images, though collected from public space, often reveal intimate aspects of everyday life. As a response, both platforms implement automated anonymization features as face and license plate blurring, in Figure 2.5. Additionally, users are granted the right to request removal of specific images, particularly when the content is considered sensitive or privacy-infringing. However, these safeguards do not fully eliminate the concerns raised by the persistent visibility and algorithmic indexing of individuals in publicly accessible visual archives.

At the same time, the enormous repositories generated by computer vision systems are becoming more permanent. The ability to withdraw from visibility, or to remain unregistered by these systems, is steadily diminishing. Such a condition undermines the possibility of a *right to be forgotten*. Control over representation thus shifts away from human intention and moves into automated, opaque processes that are rarely subject to scrutiny. In such an environment, cartographic anxiety intensifies in two ways. Ethically, the capacity to refuse participation in algorithmic classification systems is reduced. Epistemologically, the reliability and inclusivity of the knowledge these systems produce become increasingly uncertain. The ideological and representational problems long associated with cartography are therefore magnified in the



Figure 2.5: Anonymization in mapping platforms. Left: blurred faces in Karaköy. Right: obscured license plate. These show how platforms address privacy.

context of computer vision, where visual infrastructures entail far-reaching political and ethical consequences. Age of AI, where technical infrastructures are imbued with profound political and ethical stakes.

Data collected from individuals is stored indefinitely and interpreted through probabilistic models of risk. As Crampton (2011) observes, once data are gathered, they are frequently sorted into pre-existing social categories, each of which carries an assigned risk level. Populations may be marked as high-risk not due to individual behavior but because of group affiliation. Such algorithmic frameworks reinforce the logic of racial bias and sustain other entrenched forms of structural inequality. False positives and algorithmic stereotyping become normalized consequences within these systems.

At this stage, cartographic anxiety demands reconsideration. It no longer concerns only whether maps reproduce power or whether spatial truth can be objectively captured. A more existential dimension emerges: *can one escape the map?* As data accumulation becomes continuous and life-logging technologies proliferate, the ability to forget or to remain unrepresented gradually disappears. Dodge and Kitchin (2007) refers to the refusal to be continuously monitored and stored as digital information as the ethics of forgetting, calling for resistance to the totalizing capture of human life in machine memory.

In this context, mapping evolves into more than an epistemological question; it becomes

an ontological and ethical dilemma. While historically framed as a politicized act, representation is now often embedded within systems that operate as infrastructural and pervasive. As geoslavery, predictive policing, and life-logging converge, the anxiety shifts from questions of power embedded in maps to the growing sense that *unmapping the self* may no longer be possible.

What remains insufficiently explored is how states might redirect the development of AI away from harmful automation and toward more equitable and democratic outcomes. Acemoglu (2021)[7] proposes a fundamental change that involves abandoning distorted policy frameworks and revitalizing democratic governance. Without such reforms, machine learning continues to consolidate authority in the hands of those least affected by its consequences. The continued expansion of centralized algorithmic systems raises significant concerns, particularly regarding their potential social and epistemological consequences. As visual perception becomes increasingly mediated by machine processes, new modalities of knowledge production begin to take shape. In the context of this shifting epistemic condition, algorithmic systems gradually shape both the actions taken and the definitions of what is considered valid knowledge. Data that do not conform to dominant classifications or training protocols, such as those considered untrained or unlabeled, are frequently excluded from the representation. As machine perception becomes the normative mode of observation, alternative ways of knowing and inhabiting space risk being rendered illegible. In such a configuration, AI systems do not simply automate decisions. Algorithmic systems of this nature gradually disrupt the epistemic foundations that structure the way truth and visibility are defined.

In conclusion, the increasing use of algorithmic mapping and machine vision suggests a shift in cartography from a focus on representation toward more operational forms. Technologies such as facial recognition and object detection reshape the way public space is experienced. They often reinforce existing social biases and reduce the ability of individuals to control how they are seen. In the context of AI-driven spatial technologies, the idea of cartographic anxiety takes on a new meaning. It includes concerns not only about power and representation, but also

[7] Acemoglu (2021) argues that algorithmic centralization can be mitigated through a reconfiguration of democratic institutions that prioritize human-centered design over profit-maximizing automation. Without such intervention, the development of AI continues to reinforce inequality.

about ethics and exclusion from knowledge. AI now plays a role in deciding what becomes visible and how space is interpreted. Understanding how representation and technology jointly shape the epistemological and political contours of vision-based systems is essential for any critical engagement with spatial analysis today. What is outlined here offers a foundation for such an inquiry.

2.2.1 Cartographic Semiosis: Maps as Sign System

Human beings often do not engage with reality in its raw form. Instead, they tend to interpret and organize space through cultural and cognitive frameworks. In doing so, humans make sense of an otherwise complex and cognitively overwhelming world while simultaneously encoding space into a cultural meaning-making process. This coding is not purely technical; it is related to context. Symbols let us draw a quick mental sketch, communicate in a shared language, and more subtly naturalize specific socio-political settings. As Wood (1992) states, maps do not mirror the world; they create it through *a value system*. He therefore describes textit maps as semiological systems, textual fields whose signs operate on *denotative, connotative and mythic planes*.

This understanding of signs and their layered meanings draws from semiological theory as developed by Ferdinand de Saussure and later reconfigured by Roland Barthes (Barthes, 2015). Saussure proposed a dyadic model of the sign, composed of a signifier (the form of a word, sound, or image) and a signified (the concept it denotes). Barthes builds on this structure but takes it further, extending it beyond just language into the broader world of culture and ideology. At the most basic level, the denotative level, a mark on a map such as a thin black line, acts as a signifier (Barthes, 2015). At the most basic level, the denotative level, a mark on a map, such as the outline of an irregular shape, represents a lake (Wood, 1992). This contour acts as a signifier, pointing directly to a recognizable geographic feature in the real world. A shaded fill within the same shape might further specify the presence of water, reinforcing its identity as a body of water in Figure 2.6. But the process does not end there. Barthes refers to this process as *myth*, which he defines as "*a type of speech*", by which he means not spoken

language but any form of meaningful expression. This can be visual, textual, or spatial, as long as it carries communicative value. In this sense, myth belongs to the field of general semiology and operates as a second-order sign system. The Figure 2.7 illustrates this structure: at the first level, language functions through semiotics, where a signifier (such as a word or image) is linked to a signified (the concept it denotes), forming a complete sign. In myth, this completed sign is lifted into a second system and reused as a new signifier. At this second level, additional ideological meanings are attached. Importantly, the original meaning is not erased but repositioned, allowing culturally and historically specific content to appear natural and self-evident. Barthes (2015) famously describes this as the transformation of history into nature. For instance, an image of a soldier saluting a flag may at first denote patriotism, but when recoded within myth, it may connote national unity or unquestioned loyalty. Myth, therefore, is not a falsehood but a system of meaning-making that conceals its own construction. It works by appropriating familiar signs and loading them with new ideological content, all while presenting these meanings as natural and inevitable.

Building on Roland Barthes's semiological framework, Wood (1992) conceives the map not as a passive mirror of spatial reality but as a *supersign*, a synthesis of signs in which each element is layered with ideological charge in Figure 2.7. In this view, maps are systems of meaning, where signs work together to create layers of understanding. These signs combine to shape connotative and even mythic narratives that make certain ways of seeing the world seem natural and unquestionable. The semiological approach to mapping inspired by Barthes's theory of myth encourages a rethinking of cartography as an apparatus of signification, in which meaning is carefully manufactured through nested layers of semiotic encoding (Barthes, 2015).

Building on Barthes's theory of second-order signification, Wood proposes a four-layer model of cartographic semiosis—elemental, systemic, synthesis, and presentational—to account for how maps generate layered meanings. At the elemental level, basic graphic forms such as lines, colors, and symbols serve as the raw visual materials of mapping, whose meanings emerge not inherently but through cultural convention (Barthes, 2015; Wood, 1992). These elements are structured at the systemic level into codified symbol sets, forming an internal

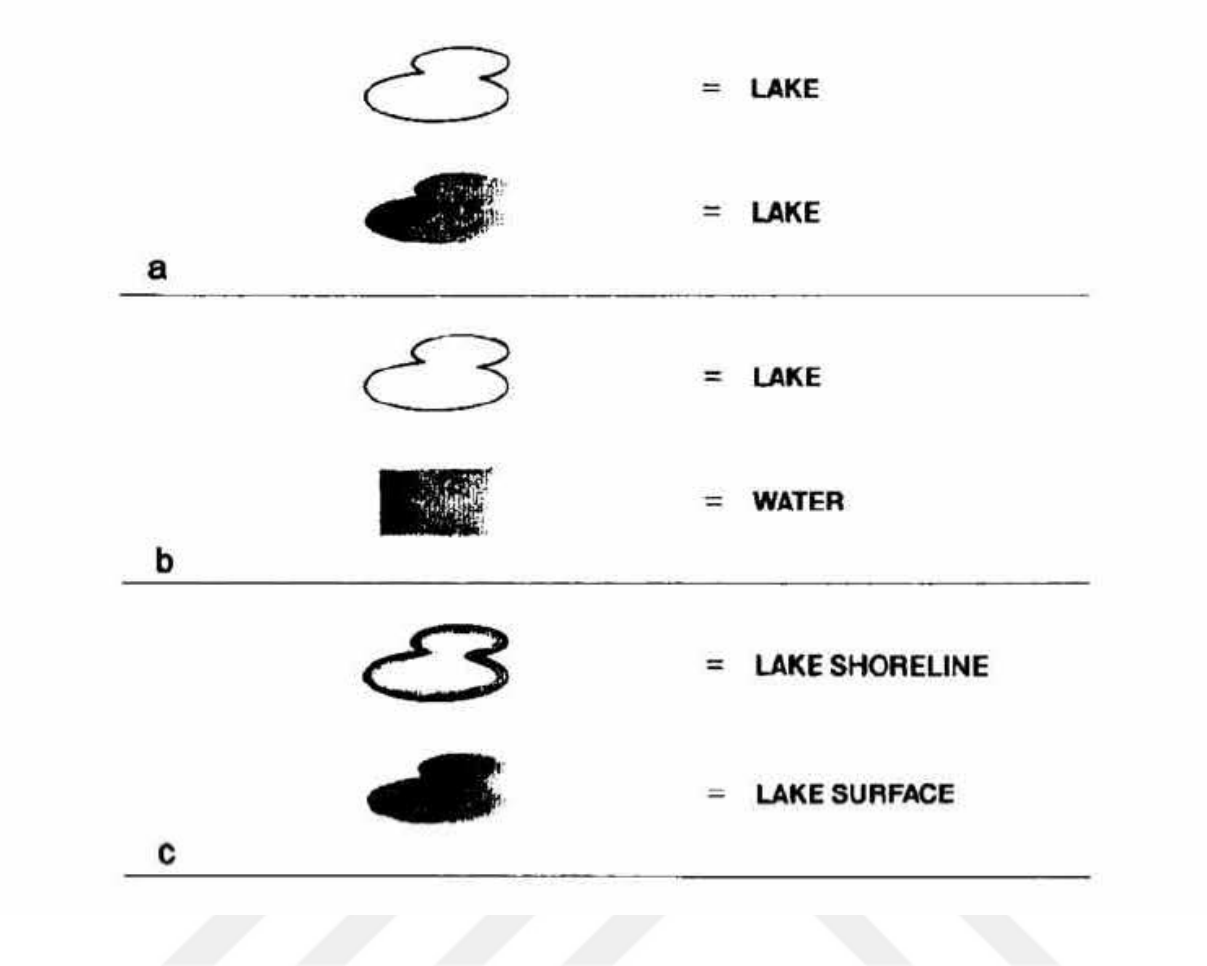


Figure 2.6: This figure illustrates that the cartographic symbol for a lake derives its meaning not from individual elements like lines or fills alone, but from their combined and context-dependent relationship as a culturally learned sign construct (Wood, 1992)

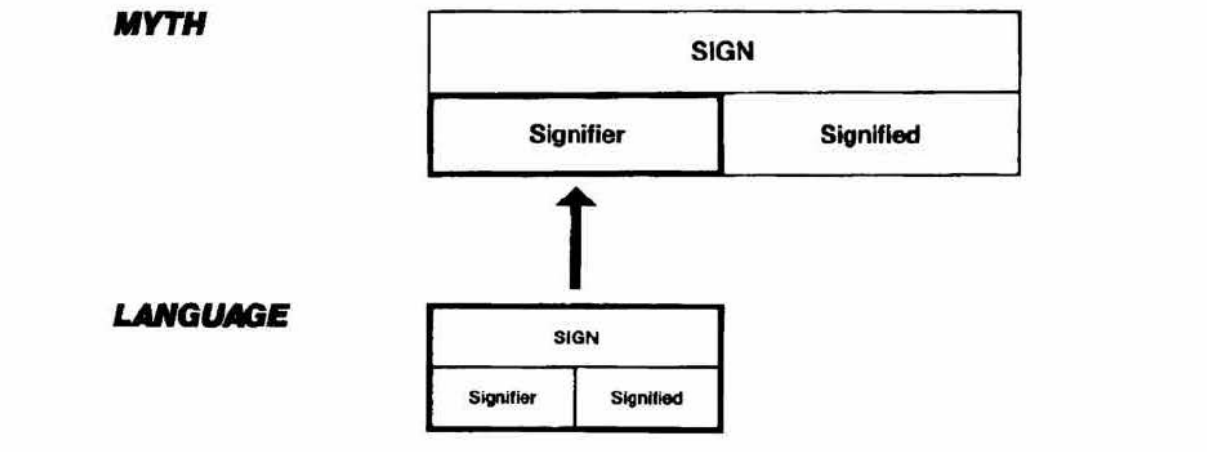


Figure 2.7: In Barthes's view, the signifier and signified form a sign, which myth then uses as a new signifier in a second-order system. He notes that describing this process as spatial is only a metaphor (Wood, 1992), redrawn from (Barthes, 2015).

grammar that governs their interpretation. The synthesis level refers to the map's overall composition, where these symbols are combined to produce spatial narratives that reflect particular viewpoints or agendas. Finally, at the presentational level, framing devices such as titles, legends, scale bars, and color hierarchies further mediate the viewer's understanding, subtly guiding perception and reinforcing the map's authority. Through these nested semiotic layers, maps operate not merely as instruments of spatial reference, but as meaning-producing structures that encode ideological assumptions, authorize specific regimes of visuality, and naturalize historically contingent worldviews as geographic common sense (Wood, 1992).

In conclusion, cartographic representations function as semiotic systems where space is shaped through meaning-making processes, rather than being passively depicted. Cultural meanings are communicated through signs and symbols, and particular ideas about space are presented as natural. The process of layered signification, involving visual elements and interpretive associations, is recognized as part of a general system of meaning-making. As Barthes (2015) and Wood (1992) conceptualize it, urban environments and architectural forms can be read as sign systems. Streets, buildings, and plans are interpreted as carriers of social and political meanings. By applying a semiotic approach, the organization of cities can be analyzed, the visibility or invisibility of certain spaces can be questioned, and the production of spatial narratives can be examined. Such an analytical approach is considered useful for interpreting existing urban systems and for designing new ones with increased awareness of the messages being communicated.

2.2.2 Computational Epistemologies of the City: From Observation to Operation

Before modern cartographic reasoning, spatial knowledge was based on everyday experience, oral stories, and performative acts like gestures, rituals, and songs. As Pickles (2004) explains, many pre-modern and indigenous societies did not measure space with abstract coordinates. Instead, these cultural groups understood space through relationships and memory. Places were remembered and navigated through stories and bodily movement, not through maps

or geometric grids. Institutional forms of knowledge production continued to shape understandings of space throughout the nineteenth and early twentieth centuries. During this period, spatial knowledge was treated as something observable from a distance, recordable on paper, and organizable into fixed categories. Tools like street atlases, social-survey maps, and master plans depicted cities as legible systems, divided into zones that could be counted, compared, coded, and brought under control (Pickles, 2004).

By the mid twentieth century, such inclusive views had begun to lose their authority. The figure of the *flâneur*, in the writings of Baudelaire and Benjamin and later the Situationist *dérive* proposed by Guy Debord, rejected the idea that cities could be fully mapped or systematized (Graafland, 2010). These forms of exploration emphasized subjective movement and personal encounter rather than formal boundaries. Collage-like maps emerged from these practices, charting informal elements such as desire paths, ambient light, or passing shadows (Crampton, 2010). Michel de Certeau argued that conventional maps erase such lived dynamics, converting the continuous flow of daily life into isolated points on a page (Crampton, 2011). Although these psychogeographic practices allowed for alternate interpretations of space, the practices rarely translated into operational strategies for managing or shaping urban environments.

Conceptions of geographic knowledge were profoundly reshaped by developments in information theory and symbolic logic. Most digital cartographic workflows are still governed by Shannon's signal-versus-noise logic, whereby information is equated with statistical regularity(Crampton, 2011). In 1948, Claude Shannon demonstrated that information could be rendered discrete, encoded into binary digits, and efficiently transmitted through communication networks (Crampton, 2011; Shannon, 1948). This epistemological shift built upon earlier foundations: Leibnitz (1703) had shown that knowledge could be represented through binary structures (0 and 1); Boole (1847) formalized these representations through logical operations and algebraic rules; and Shannon finally proved that such logic could be physically processed by digital machines, transforming abstract information into electrical signals and enabling new modes of spatial abstraction and transmission in Figure 2.8. Influenced by Shannon, mapping was reconceptualized not just as *depiction but as a form of transmission*, emphasizing

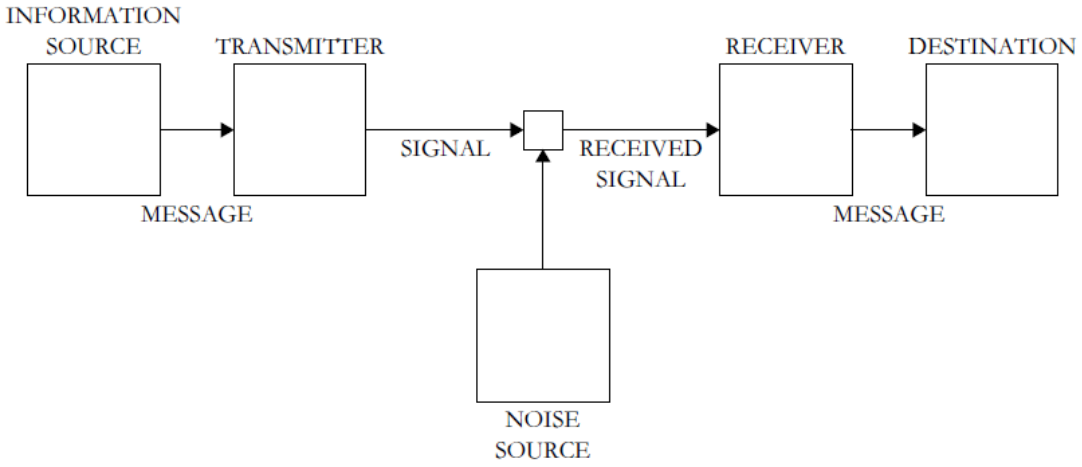


Figure 2.8: Shannon's 1948 schematic of communication system, redrawn by Crampton.

informational clarity and the minimization of uncertainty over traditional representational accuracy. Here, the critical issue lies in the *fungibility and computability of information*. Drawing upon Claude Shannon's information theory, knowledge today is reduced to its smallest possible units—binary bits of zeros and ones, yes and no—making it transferable, calculable, and operational. Consequently, information has ceased to be simply representative or observational; it now functions as an actionable entity processed by algorithms and machines. With the emergence of automated processing and machine learning, the epistemology and ontology of spatial knowledge have changed. According to Foucault (2013), the surface of knowledge displays a discontinuity: history, once imagined as a continuous and linear narrative, now appears fragmented, constituted by discontinuities and the emergence of new epistemic ruptures. Today's digital knowledge surface, therefore, differs fundamentally from previous representational practices; data often conforms to the requirements of automated systems and algorithmic routines, and observational traces, once passively recorded, now act as active agents that directly shape interventions in the physical world.

However, the noise of Shannon's theory could serve as *critical evidence* in cartography. Malfunctions, coverage gaps, and signal losses are not simply technical failures; they often reveal patterns of infrastructural exclusion or neglect. A critical cartographic approach takes such irregularities seriously, reading *them as symptomatic rather than accidental*. By doing so,

it expands the conceptual boundaries of mapping, allowing silence, absence, or distortion to function as meaningful indicators of how urban space is governed and inhabited.

Shannon's model, while foundational, remains confined to a syntactic domain—it concerns the form and fidelity of transmitted signals, not their significance. In response, Sequoiah-Grayson and Floridi (2022) advances a semantic model of information: for data to count as information, it must be both meaningful and factive. This shift introduces a normative filter—correcting Shannon's reductionism by grounding information in the domains of meaning and epistemic validity.

However, Floridi's model also imposes a new limitation. By prioritizing truth, it risks reinforcing a rationalist epistemology that can obscure the contingent, interpretive, and ideological dimensions of meaning (Sequoiah-Grayson & Floridi, 2022). The question, then, is not simply whether a pattern is statistically regular or semantically valid—but what that pattern means, why it appears, and whose interests it serves. A wisp of smoke may suggest fire, but the inference must be examined semantically: under what conditions is that link persuasive and for whom?

Semiology insists that meaning is never self-evident. It emerges within networks of codes, conventions, and cultural matrices (Barthes, 2015). A photograph of a factory may denote industrial activity; symbolically, however, it may connote power, productivity, environmental degradation—or all three—depending on the discursive context in which it circulates. Logical truth conditions alone cannot exhaust these layers of signification.

The limitations of a purely linguistic model become even more pronounced in spatial analysis. Lefebvre (2012) asserts that space is not a neutral container awaiting representation, but an ongoing product of social and symbolic practices. Traditional cartographic models flatten this dynamism, translating space into a legible surface to be read and decoded. This representational act erases the very processes by which space is shaped—through power, experience, and struggle. To read space semiologically is to recognize that it is not merely mapped but performed, contested, and rendered meaningful through signs. Mapping, then, is never neutral; it is an operation of control that simultaneously reflects and reproduces the structures of power

it appears merely to depict.

This tension between representation and reduction continues to haunt spatial knowledge production. Even within more advanced models such as Sequoiah-Grayson and Floridi (2022), the transformation of space into "information" remains problematic. As Henri Lefebvre emphasizes, the very act of converting space into knowledge is itself a form of reduction. Spatial knowledge, he argues, is never merely retrieved or recorded—it is produced through embodied, symbolic, and contested practices. No informational model—whether syntactic, semantic, or symbolic—can fully capture the complexity of space. This epistemological concern has not been resolved with the digitality of the systems; on the contrary, it has intensified with the advent of algorithmic governance and artificial intelligence. As data infrastructures become more pervasive and automated, the production of spatial knowledge is increasingly mediated by computational systems that privilege operational legibility over interpretive depth.

What emerges from the epistemological trajectory outlined above is a critical awareness of how digital systems, particularly GIS platforms, remote sensing technologies, and visual infrastructure like Street View, reconfigure the ontology of space. Rather than solely observing space, they frequently participate in its operationalization. Mapping becomes an act of computational reduction, in which spatial reality is parsed into machine-readable codes and filtered through predefined ontologies. Consequently, every spatial decision, from labeling a road to identifying a facade transformation, is mediated by algorithmic routines that embed both technical and ideological assumptions. In the context of this thesis, which utilizes machine vision and structured visual interpretation to detect micro-scale changes across time. The critical awareness that digital spatial systems remind us that such visual outputs are not passive representations but products of a data pipeline built on Shannon's logic of transmission and coupled with GIS-driven classification schemas. Therefore, in producing spatial knowledge through these tools, one must remain attuned to their reductive tendencies. The goal is not to reject operational mapping but to engage with it critically, recognizing where it reveals structure and where it silences meaning.

2.2.3 Machine Semiosis: Algorithmic Reading and Visual Parsing

Space is no longer apprehended as a purely physical or experiential entity; it is often conceived as a domain systematically transformed into distinct types of information. As Adriaans (2021) outlines, information can be categorized into three forms: *Information-A*, referring to meaningful, logically structured content comprehensible to human cognition; *Information-B*, derived from Shannon's model, which reduces signals to statistical units measured by probability and entropy; and *Information-C*, which concerns algorithmic compressibility and computational complexity. Through digital representation, spatial environments are increasingly reconstituted within these informational frameworks. Visual elements such as images may be converted into tensors, colors into numerical matrices, and sensory experience into RGB channels. In this process, space does not necessarily persist in its full phenomenological form but is often reduced, at least in operational terms, to *the mode of its representation*.

Neural networks may exemplify the statistical abstraction. They operate by relating features not through meaning, but through statistical proximity: attention mechanisms identify which signals co-occur frequently, but not why they matter. The internal architecture of neural networks is built from previous data associations; space becomes a matrix of values, where each encoded layer corresponds not to lived depth but to algorithmic salience. Generative Adversarial Networks (GANs), while capable of producing photorealistic imagery, function only within the logic of singular frames and fixed perspectives.[8] Their outputs simulate realism but never reconstruct the multiplicity of a lived spatial encounter. As the outputs of GANs, they reflect only one sensory modality, namely vision, and fail to approximate the richness of a space that is also touched, heard, smelled, and embodied.

What results from the spatial reduction is not spatial knowledge in the epistemic sense, but a relational encoding, a form of mapped cognition, where numerical proximity substitutes for semantic interpretation. The process entails loss at every stage: what is encoded is always

[8]Generative Adversarial Networks (GANs) are artificial intelligence models designed to generate realistic content by learning from real examples. They consist of two components that work in opposition: one generates new outputs, such as images, while the other evaluates them and provides feedback. Over time, this dynamic helps the model produce outputs that closely resemble real-world data.

less than what is lived. Each translation involves abstraction, omission, and flattening. Even the RGB values used to express spatial textures participate in a hierarchy of attention: low-level features like color gradients are transformed into high-level features like object labels, yet the structure of these mappings is governed entirely by pre-trained associations (Newman, 2024).

The reduction brings into focus the epistemological opacity of such processes. As Harley (1989) contends, mapping has become *a black box*—a system whose operations, assumptions, and exclusions are inaccessible to the very actors who rely on its outputs. The same may apply to artificial intelligence models. Although their results can be interpreted, the processes by which those results are produced are rarely understood. Transformer-based attention mechanisms are often praised for their accuracy, but what is produced is not meaning. It is a statistically optimized correlation. [9]

In this sense, *attention itself becomes a form of mapping*, yet one that encodes without disclosing. The ontology of space is thereby restructured: it is no longer lived, but parsed, no longer embodied but calculated. In doing so, it establishes what is important according to its internal logic, thereby also determining what qualifies as *"information"*.

This transformation is grounded in the principle of *relationality*. As Acar (2019) argues, mapping is essentially an act of linking, comparing, and organizing; it often constructs the structural and semantic relationships through which knowledge becomes legible. No informational unit stands alone; it acquires meaning only in context, through association. Patterns are reinforced not because they hold intrinsic meaning, but because they appear with sufficient frequency to be recognized as significant by the model. What is prioritized is not semantic relevance, but statistical consistency. These systems do not evaluate what matters in a human sense; rather, they are calibrated to detect and amplify recurrence. Accordingly, AI outputs are not epistemically neutral, as they are shaped by the assumptions embedded in the training data and the design choices underlying the model (Newman, 2024; Vaswani et al., 2017).

[9] Transformer-based attention mechanisms are techniques used in artificial intelligence to help the model focus on the most relevant parts of the input. The term *"attention"* refers to the model's ability to weigh different pieces of information and decide which ones are more important for the task, such as identifying keywords in a sentence or important areas in an image (Vaswani et al., 2017).

Computer vision refers to a subfield of artificial intelligence that enables machines to interpret and extract information from visual data, such as images or videos. It typically involves detecting patterns, segmenting objects, labeling features, and generating representations based on pixel-level input. However, beyond its technical aims, the process of visual interpretation performed by such systems invites broader reflection on questions of signification and meaning. In this context, what might be termed ”machine semiosis” offers a useful conceptual lens. Machine semiosis refers to the idea that algorithmic systems may not merely record visual input but instead participate in the active production of signs. When a convolutional network (CNN) processes an urban streetscape, it begins by extracting numeric features—edges, gradients, and histograms. These raw values are quickly inscribed with semantic labels such as “graffiti,” “window,” or “vacant lot.” This two-stage translation mirrors classical semiotics: a *denotative* layer of measurable form is overlaid by a *connotative* layer of meaning, itself conditioned by prior datasets As a result, the algorithm’s gaze is never neutral; it foregrounds the visible where examples are abundant and renders invisible the different, the unstructured or the undocumented in Figure 2.9.[10]

The dashboards and maps generated downstream inherit these visual hierarchies, reinforcing them through the illusion of neutrality. Recognizing this dynamic expands critical cartography into the domain of AI. It requires us to ask not only what is being represented, but how and why such representations are constructed. Auditing datasets, making architectural assumptions explicit, and integrating plural epistemologies become essential to producing accountable spatial analytics. The task is not to abandon automation, but to situate it within the politics of representation—transforming machine vision from an inscrutable oracle into a reflexive partner in understanding the complexities of urban life.

Machine vision in contemporary AI does not simply capture spatial reality; it may actively construct it through *mechanisms of attention*, statistical association, and inherited classifications. Rather than producing spatial knowledge in a phenomenological sense, it generates

[10]The second example includes a carved niche resembling a religious iconography. Although structurally analogous to standard windows, this culturally loaded form does not elicit the same attention response. The model seems to respond based on both the shape of the object and how commonly similar forms occur in the training set. In this case, unfamiliar aesthetic or symbolic motifs are given less emphasis, which reflects a form of dataset bias present in the model’s learned gaze.

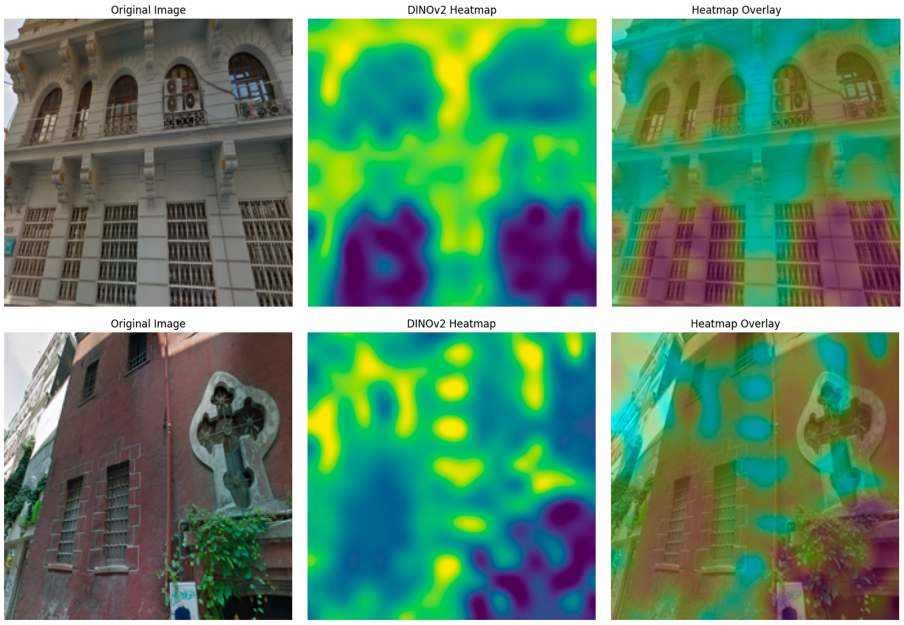


Figure 2.9: Heatmap comparison between two urban scenes as interpreted by DINOv2. In the first row, the model clearly attends to familiar architectural features such as window bars and facade symmetry. In the second row, although there is a visible opening similar to a window, the model's attention is less focused and more dispersed.

11

computational abstractions shaped by prior data. This critical view of machine vision may invite a reconsideration of how algorithmic systems render space intelligible and to what extent such renderings can be treated as meaningful reflections of urban reality. What becomes visible, what remains obscure, and how these distinctions are constructed will remain open questions throughout the analytical process.

Street View imagery has transitioned from a centralized visualization tool to a distributed infrastructure for spatial data production. Its expansion through diverse capture methods and participatory platforms reflects a broader shift toward operational and algorithmic modes of mapping. Yet, this apparent democratization raises critical concerns regarding authorship, labor, and representational politics. As visual geographies become increasingly shaped by platform-mediated systems, ensuring transparency, equity, and accountability in their design and governance remains essential.

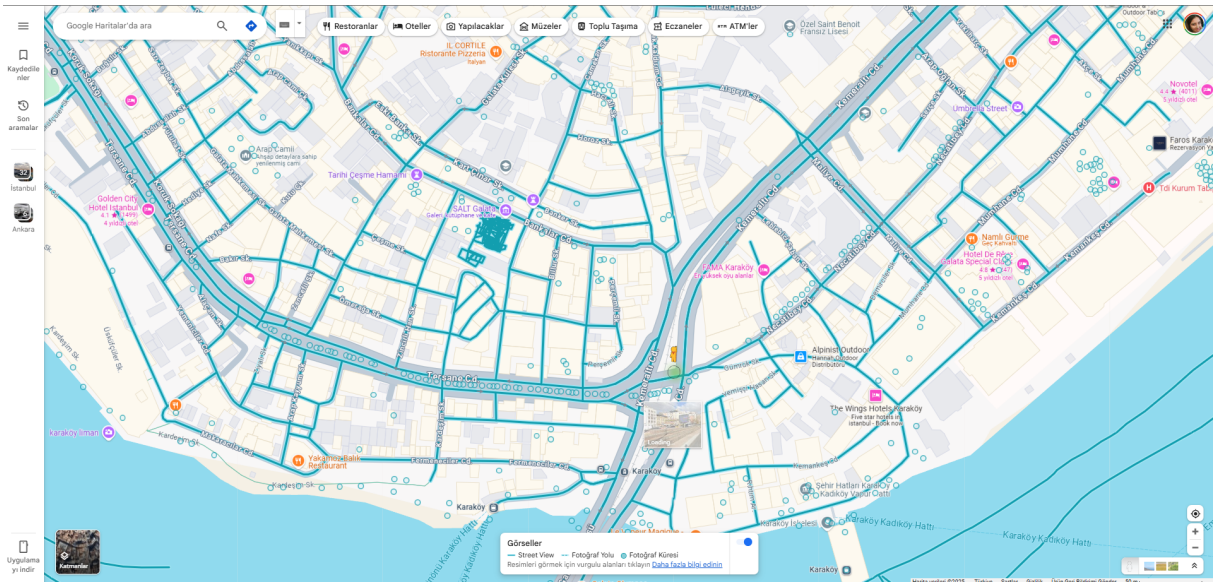


Figure 2.10: A Google Maps interface displaying blue-highlighted roads, indicating Street View availability, with Pegman ready to be placed.

2.3 Street View Imagery as a Tool for Urban Operation

Although much of the discussion around GSV has focused on its technical sophistication and global reach, the way users encounter and navigate the platform in everyday life owes much to its deliberately intuitive interface. Integrated into Google Maps, Street View invites interaction through a familiar gesture: dragging the small yellow figure known as *Pegman* onto blue-highlighted streets in Figure 2.10. From that point, users are immersed in a 360-degree visual environment where they can look around, zoom in on urban and building details, and move forward or backward along the street with a click—mimicking the experience of walking through a city in Figure 2.11. Such interaction, while simple on the surface, reshapes how people understand and connect with place: not through abstract coordinates but through images charged with presence, memory, and affect. It is precisely this capacity to simulate embodied spatial exploration that has made Street View feel seamlessly woven into everyday urban experience (Gilge, 2016).

The first version of Street View originated through a collaboration between Google and the Stanford Computer Graphics Laboratory, where efforts were directed toward developing



Figure 2.11: The Street View interface as accessed through Google Maps

new methods for representing the experience of moving through urban space. This initiative culminated in the CityBlock Project, an early attempt to generate panoramic imagery capable of summarizing entire city blocks within a single continuous view (Anguelov et al., 2010; "Stanford CityBlock Project", 2013). Using a sideways-facing high-speed video camera on a slow-moving vehicle, the Stanford team developed a technique called multi-perspective panoramas. By extracting narrow vertical slices from sequential video frames and combining different types of projection, they produced fluid images of streetscapes with minimal distortion ("Stanford CityBlock Project", 2013). Although this *strip-based method* was not used in the final version of Street View, it helped shape the project's early direction. [12]

The CityBlock Project introduced *the idea of crawling the physical web*, a phrase developed with Larry Page to describe the large-scale, image-driven mapping of urban environments. The early concept provided both the technical base and the vision for what Street View would eventually become: a global platform for navigating the visible world ("Stanford CityBlock Project", 2013).

[12]The strip-based method refers to an approach in which sequences of images are captured along a linear path—such as a street—then algorithmically stitched together to form continuous panoramic representations of entire blocks or routes. Unlike isolated panoramic nodes, this method emphasized spatial continuity and narrative coherence along a trajectory.



Figure 2.12: A Google Street View car equipped with a panoramic camera system used to capture 360-degree imagery for Google Maps (Google, n.d.).

Street View, as it was eventually deployed, shifted from *strip-based mosaics to discrete 360-degree* panoramas captured at intervals along streets. Following this shift, Google began building custom imaging platforms including rooftop-mounted panoramic camera rigs (Anguelov et al., 2010; WIRED, 2022) (Figure 2.12). Early systems such as the R2 and R5 platforms were capable of capturing high-resolution imagery and depth data simultaneously, enabling the construction of photorealistic street-level views with millimeter-level geometric accuracy (WIRED, 2022).

Over time, Google Street View (GSV) was gradually enhanced through the incorporation of advanced image processing pipelines. Tasks such as data anonymization, geospatial alignment, and 3D model fusion were implemented to support features including virtual navigation and enhanced panoramas (Anguelov et al., 2010). In more recent developments, compact and modular camera systems were introduced, allowing data to be collected using a wide variety of vehicles. Additionally, the platform was extended to include immersive visualization tools such as Immersive View, which integrates street-level imagery with aerial photographs to simulate urban environments in real time (WIRED, 2022).

The expansion of image capture methods was accompanied by a significant diversification in mobile data collection apparatuses. In addition to vehicles, tricycles were deployed in



Figure 2.13: Examples of specialized Street View capture devices adapted for diverse terrains, including a tricycle (2008), snowmobile (2010), and Trekker backpack (2012) (WIRED, 2022).



Figure 2.14: Alternative mobile platforms used in Street View projects, showing a camel-mounted camera in the Liwa Desert and a sheep-mounted system in the Faroe Islands (Parkinson, 2014; Vega, 2016).

heritage zones with narrow streets, trolleys were used in interior spaces such as museums and stadiums, snowmobiles were adapted for alpine environments, and backpack-mounted Trekkers were carried along wilderness trails, as illustrated in Figure 2.13. The expansion of image capture methods was further exemplified by the 2016 *Sheep View 360* initiative in the Faroe Islands and *Camel View* in Liwa desert in Figure 2.14 (Parkinson, 2014; Vega, 2016). In that sheep project, 360-degree solar-powered cameras were attached to roaming sheep by local tourism officials in order to map remote cliffside paths inaccessible to motorized vehicles. As a result of this initiative, additional trekker kits were provided by Google to support local engagement. Since then, the range of capture methods has continued to expand, with devices now being mounted on bicycles, kayaks, wheelbarrows, and even horses. In this way, any mobile unit has been rendered capable of contributing to the global imaging system (Vega, 2016).

Through such efforts, the boundary between observer and instrument increasingly blurs. Any mobile entity—whether human, animal, or mechanical—functions as *a data-gathering apparatus* by virtue of its ability to move through and experience space (Vega, 2016). As a result, the act of traversing physical environments becomes a process of generating visual information and encoding spatial data within the wider Street View infrastructure. While this dynamic was initially developed and centralized through corporate platforms such as GSV, similar logic has since extended into more decentralized and participatory systems. Crowdsourced imagery platforms have emerged as alternatives, enlisting individuals as mobile sensing agents to contribute to shared visual mapping efforts. Among the most notable examples is Mapillary. [13]

2.3.1 The Myth of Neutrality in Street View Imagery

At first glance, GSV appears to represent the street-level environment as it truly exists. In truth, it offers selected representations that emerge from a multilayered process of editing and encoding. Just as satellite images captured at different wavelengths are later recolored to normalize them, GSV imagery is reshaped through automatic color correction, blurring and even manual intervention, which are rooted in assumptions embedded in software, sensors and algorithms (Anguelov et al., 2010).[14] Rather than reflecting pure reality, GSV functions as a visual map constructed through specific coding choices and framing strategies. A sign system produces myth only when situated within a cultural structure, so the photograph itself need not contain explicit relays of meaning (Barthes, 2015).[15] Although photography is often reduced to its perceived realism, GSV's layered production conditions create new possibilities for myth-making. As Lefebvre (2012) emphasizes, all visual representations—be they maps or photographs—inevitably entail selective omission, foregrounding certain spatial elements while

[13]Founded on open-source principles, Mapillary enables users to upload geotagged street-level imagery using everyday devices such as smartphones and dashcams. Positioned as a bottom-up alternative to centralized platforms, it aims to increase coverage in areas underserved by services like Google Street View. However, despite its participatory veneer, Mapillary operates within a platform-capitalist model: it monetizes user-generated content while outsourcing the labor and risk of data collection to contributors (Biljecki & Ito, 2021; Cinnamon & Gaffney, 2022).

[14]This refers to atmospheric correction and spectral transformation techniques commonly applied to satellite imagery, such as those from Landsat or Sentinel, to render them in a more visually 'natural' color range.

[15]This reflects Barthes' theory of second-order signification, where a denotative sign becomes a connotative myth when embedded within an ideological or cultural context.

marginalizing others. As a result, each image constitutes a spatial abstraction, whose meaning is produced not only through what is rendered visible but also through what is systematically excluded.

According to Barthes (2015), when a place has not been directly encountered and firsthand relations are absent, the photograph is positioned as the primary object through which that space is accessed and thus turns into a consumable image. This process transforms the photograph from a neutral record into an object of interpretation and narration. When direct experience is replaced by mediated vision, the subject no longer walks through space but instead engages with the image of someone else performing that action. The relationship with space becomes distanced and abstracted. In such cases, the image does not merely depict an event or location; it becomes something to be described, interpreted, and consumed. Within the framework of platforms like GSV, this abstraction becomes more pronounced. Street-level imagery captures not only architectural elements or urban infrastructure but also anonymous individuals caught in the middle of everyday gestures. These individuals appear as graphic elements rather than as agents in their own spatial stories, their presence flattened into signs and their movements stripped of context (Barthes, 2015).

In the context of GSV, this reduction becomes more pronounced. Bodily experience; wayfinding, observation, and spatial exploration are replaced by an interface addressed solely to the visual senses. Platforms such as GSV function as surrogates that supply synthetic visual layers in place of embodied spatial experience. While Descartes imagined the subject as a rational mind navigating an abstract ordered space, the embodied subject is located in a world structured through practice (Gilge, 2016). Merleau-Ponty's theory emphasizes that the body is not simply a perceiver but a condition for perception; space is apprehended through movement from a specific bodily origin and directionality as well as spatial meaning are relational and tied to corporeal positioning (Gilge, 2016; Merleau-Ponty & Smith, 2015).[16] Because GSV fixes the viewpoint and removes directional cues (for instance, the loss of a meaningful north arrow), these embodied references fade and their experiential value declines.

[16]Merleau-Ponty challenges the Cartesian model of the disembodied rational subject, arguing that spatial understanding is always grounded in bodily orientation and movement (Merleau-Ponty & Smith, 2015).

What emerges is a form of synthetic image construction that provides a visual stand-in for physical presence. *Situatedness*, once grounded in lived bodily experience, becomes abstract (Gilge, 2016). Perceptual actions that once involved interactive engagement with the environment risk being reduced to passive visual reception. In his foundational work, Lynch (1964) emphasized how wayfinding is deeply embodied: orientation in space is not merely a visual function, but one constructed through observation and multi-sensory engagement. Spatial intelligibility, in Lynch's view, emerges from traversing space, from a bodily immersion in the urban environment's textures, rhythms, and cues (Lynch, 1964). However, the immersive platforms function as visual apparatuses that simulate presence while abstracting away embodied experience. They decouple navigation from the corporeal field, replacing the lived construction of spatial memory with disembedded visual sequences. These platforms do not support the experiential accumulation of spatial knowledge through movement and orientation; instead, they offer pre-encoded and fixed visual fragments that are viewed, not lived.

This disembodied visuality is not merely a technical feature of GSV, but must be understood as part of an apparatus of visual power. According to Michel Foucault, a dispositif refers to a heterogeneous ensemble composed of discursive practices, institutional mechanisms, architectural forms, regulatory norms, and modes of knowledge that work together to guide, structure, and govern both subjects and spaces (Foucault, 2020). Foucault offers several emblematic examples of dispositifs: the prison, the clinic, the confession, and sexuality, each demonstrating how power operates not through coercion alone but through the structuring of visibility, normativity, and knowledge (Foucault, 2020).

GSV could be analyzed as a contemporary visual dispositif that synthesizes multiple layers: technical infrastructure (panoramic imaging, data APIs), institutional control (Google's platform), discursive authority (the privileging of visual realism as spatial truth), and subject formation (the passive, observing user). GSV disciplines space by regulating what can be seen, framing urban environments through a pre-coded, algorithmic gaze. It replaces embodied navigation with image consumption, producing a disembodied subject who experiences space through visual substitution rather than movement.

Furthermore, GSV does not represent all spaces equally: highly visible zones tend to align with commercial, touristic, or high-investment areas, while marginalized or vacant sites are often excluded. This selective visibility reinforces spatial hierarchies, rendering some areas hyper-visible while others remain unseen. As such, GSV constitutes not merely a technological tool, but—following Foucault's formulation—a dispositif that operates through vision and digital mediation to produce new configurations of urban power and perception.

In addition, GSV incorporates a layer of prosumptive participation: users not only consume but also generate content—images, reviews, ratings—which shape others' perceptions and guide spatial behavior. This input is regulated through gamified mechanisms (points, levels), forming a soft disciplinary system that incentivizes ongoing engagement. Moreover, the visibility of places correlates with zones of economic investment and service density, thereby reproducing spatial inequalities through algorithmic mediation (Foucault, 2020; Zajc, 2015).

GSV also operates through specific spatial and temporal constraints: images are taken at fixed intervals, from an elevated viewpoint, and rarely capture the layered social dynamics of the street. Street View vehicles travel along selected streets at specific times; thus, the images they capture cannot document every hour of the day or every layer of the city. The camera's height and viewing angle often differ from an everyday human perspective, and the ongoing social and cultural interactions within the streets rarely surface fully in these panoramic photographs. All these stages and decision points extend well beyond efforts to make the images technically accurate or high-resolution. They also involve political, social, and ethical dimensions, such as which regions receive more frequent updates, which details are concealed, and which are brought to the fore.

These spatial and temporal constraints are not simply technical limitations; they shape the kinds of knowledge that can be derived from Street View imagery. What is included, excluded, or only partially visible in the images directly affects how researchers interpret urban form and change. As a result, any analytical use of GSV must contend with the platform's selective ways of seeing and recording the city. The imagery archive remains curated rather than neutral, mirroring platform-driven rhythms of visibility. These representational boundaries influence

not only qualitative observations but also large-scale computer-vision analyses of urban space.

The effects of these representational constraints are especially true for studies that require large-scale data or focus on long-term trends (Curtis et al., 2013; Kim & Jang, 2023). Biljecki and Ito (2021) and Li (2021) have used GSV imagery to identify and evaluate urban features such as sidewalks, street trees, intersections, signage, and overall spatial form. Thanks to developments in computer vision and deep learning, GSV has also been integrated into automated systems that assess street greenery, estimate urban heat exposure, measure walkability, and approximate perceived street safety (Biljecki & Ito, 2021; Sun et al., 2022).

Despite these advantages, several studies have pointed out important limitations, especially concerning GSV's *spatial coverage* and *temporal reliability*. These factors are often underestimated in applied research. A main concern is the inconsistency in the time frames of GSV imagery. One of the earliest studies, conducted by Curtis et al. (2013), found that in five American cities image dates frequently shifted along short road segments, often within 25 m of intersections. These date changes are usually not visually obvious and can mislead researchers into assuming a single point in time in Figure 2.15. More recently, Kim and Jang (2023) analyzed nearly 100,000 home-to-work walking routes in 45 small and medium-sized American cities in Table 2.1. They discovered that 44% of the routes lacked adequate GSV coverage and that the mean time span along a single path was about seven years; only 10% had imagery captured within one year. Such temporal fragmentation can lead to significant measurement errors, especially in studies of environmental exposure or change.

City and Associated Counties	Population Sample	Commute Time (min)	Spatial Coverage	Season-level	Month-level
Bloomington-Normal (McLean, IL)	1763	19.99	86.4	1.1	73.8
Muncie (Delaware, IN)	1409	20.27	85.6	1.3	115.9
Bloomington (Monroe, IN)	3447	20.02	86.9	1.5	58.1
Lafayette-West Lafayette (Tippecanoe, IN)	3281	20.34	90.0	0.4	81.3
Cedar Falls-Waterloo (Black Hawk, IA)	1000	19.65	70.4	0.4	16.2
Iowa City (Johnson, IA)	4788	20.43	96.4	0.7	75.2
Ames (Story, IA)	1817	19.73	93.6	0.7	67.5
Lawrence (Douglas, KS)	1936	20.92	94.6	1.6	107.2
Manhattan (Riley, KS)	1933	19.94	94.7	1.4	108.2
Clarksville (Christian, KY; Montgomery, TN)	281	20.02	95.2	1.4	48.0

Table 2.1: Summary statistics for the first 10 selected U.S. cities (Kim & Jang, 2023)

Additionally, research shows that these inconsistencies are not randomly distributed. According to (Fry et al., 2020; Quinn & Alvarez León, 2019), areas with higher income and heavier

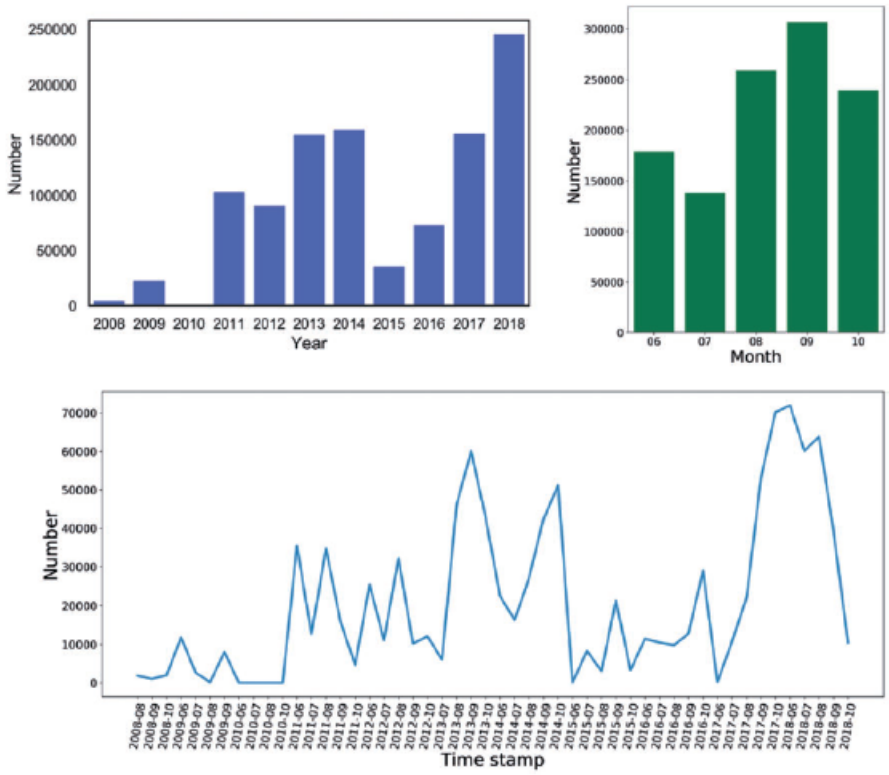


Figure 2.15: The number of GSV panoramas available for New York City, shown by year (Curtis et al., 2013).

traffic are more frequently updated and better covered, whereas lower-income neighborhoods and peripheral districts often have outdated or missing imagery, raising concerns about spatial inequality in data representation.

These empirical patterns expose methodological challenges and reveal epistemological tensions inherent in mapping technologies such as GSV. Temporal discontinuities in imagery underscore that cartographic practices are not merely spatial representations but are deeply embedded in temporal structures. By privileging spatial fixity and marginalizing temporality, such practices risk reifying the urban environment as static and timeless, thereby erasing the temporal heterogeneity of lived space. Wood (1992) problematizes this spatial bias by demonstrating how mapping compresses dynamic processes into static visual forms. His analysis of a bus journey reveals how temporally unfolding experiences are transposed into abstract spatial traces, reducing duration to mere geometry in Figure 2.16. In this context, mapping emerges as a selective operation that constructs legibility through the suppression of temporal complexity. Addressing these dynamics requires conceptual frameworks that reconceptualize mapping

as a spatio-temporal practice—one that both produces and conceals the rhythms, ruptures, and contingencies of urban life (Wood, 1992).

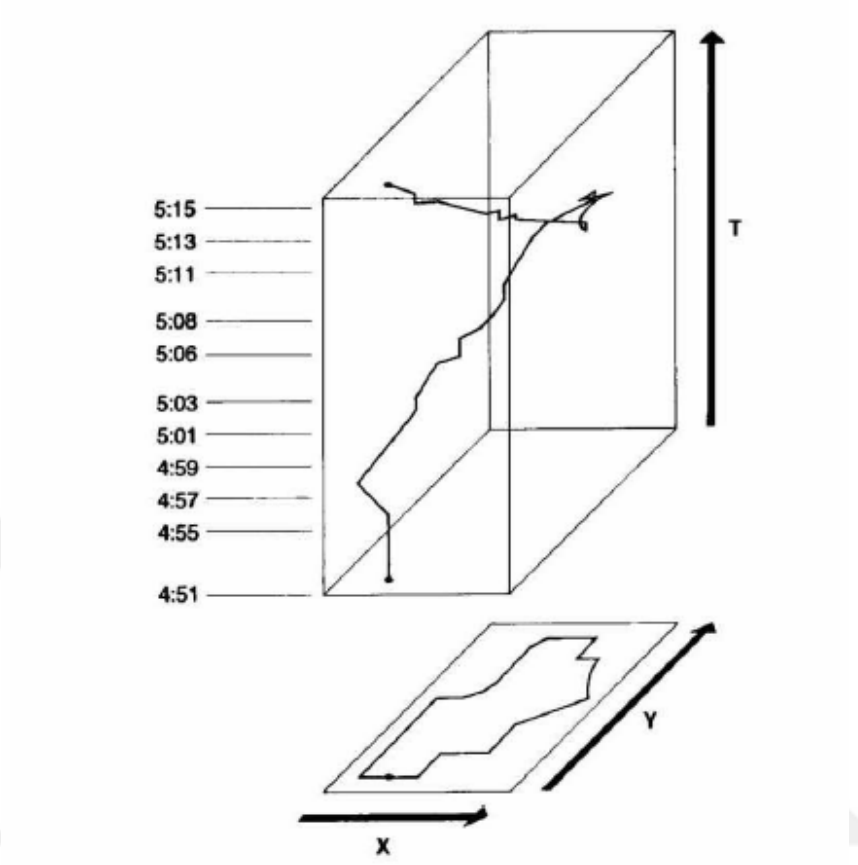


Figure 2.16: A spatio-temporal bus trip map to temporal flattening

The spatio-temporal approach to cartographic representation supports Wood's claim that maps are not neutral reflections of the world but instruments that shape perception. By eliminating time, maps conceal historical change, social dynamics, and processes of transformation. What looks like a stable and objective view of space is actually produced through deliberate representational choices that mask ongoing economic, political, and cultural forces. To explain this, Wood introduces the idea of the *durative code*. Every map carries an implicit temporal scale that links the amount of real-world time represented to the duration visible on the map. A digital animation that compresses 24 hours of vehicle movement into 1 minute illustrates how this ratio drastically shortens perceived duration (Wood, 1992). Street View vehicles operate according to a sequential recording logic; however, the resulting imagery seldom corresponds to a unified temporal layer. Even when images are captured within the same month and year, they may originate from distinct temporal contexts, reflecting discontinuities that undermine

assumptions of synchronicity and temporal coherence.

Rather than serving as a real-time or temporally neutral representation of the urban environment, GSV should be conceptualized as a durative cartographic medium that constructs spatial meaning through the selective layering, omission, and compression of time. Phenomena such as seasonal vegetation, incremental renovations, or the appearance and disappearance of signage expose latent temporalities embedded within the platform's visual interface. When approached critically, GSV emerges not as a transparent window onto urban space but as a sortable temporal archive shaped by algorithmic rhythms, update cycles, and interface constraints. These mechanisms constitute a digital enactment of the durative code, encoding temporality not as linear progression but as fragmented visual intervals that simulate continuity. Acknowledging the presence of these temporal inscriptions does not undermine the analytical value of GSV data. Rather, it necessitates a reflexive methodological stance that incorporates triangulation through complementary sources such as field observation, archival documentation, and situated local knowledge.

2.3.2 Implementations of Street View Imagery

Street-level imagery has increasingly been used as a core medium through which urban environments are observed, evaluated, and managed. It has been incorporated into a range of applications, from geospatial analysis and real estate to navigation and planning systems. While urban areas become denser and more responsive to data-driven approaches, street view imagery has been repurposed not only for orientation or visual documentation but also for analytical analysis such as perception modeling (F. Zhang et al., 2018), structural classification (Ogawa et al., 2023), energy simulation (Szcześniak et al., 2022), walkability assessment (Zhou et al., 2019), evaluation of visual aesthetics (Zhou et al., 2021), greenery measurement (Li et al., 2022), and the detection of building functions (L. Zhang et al., 2020). These developments suggest that cities are no longer simply viewed but are being computed. A shift has occurred from observational cartography to operational urbanism. Through the integration of machine-readable visual content into urban workflows, street view imagery has been transformed from

a passive visual archive into an active input for governance, equity analysis, and environmental planning.

This transformation indicates a shift in how cities are studied. Researchers now treat visual data not as secondary or decorative but as central to how urban knowledge is produced. Instead of basically looking at cities, they read, model, sort, and act on them through large-scale, image-based analysis. As computer vision tools improve and platforms like Google Street View grow, researchers apply street-level imagery to tasks such as analyzing urban form, predicting risk, and mapping inequality. In recent years, several leading urban digital labs that initially focused on mapping and platform-based research have started to incorporate street-level visual data into their investigations. Labs such as MIT's Senseable City Lab, ETH Zurich's Future Cities Lab, and the Urban Analytics Lab at the National University of Singapore have emerged as pioneers in this transition, integrating computer vision and geospatial imagery into urban research workflows (Acar, 2019; ETH Zürich, 2025; MIT Senseable City Lab, n.d.; NUS Urban Analytics Lab, n.d.). These labs use street-level imagery within broader digital systems for sensing, simulating, and planning urban environments. In these projects, the street becomes more than a visual or symbolic space.

One dimension of this framework focuses on *building-level inference*. By aligning panoramic imagery with GIS-based building footprints, researchers have developed methods to estimate architectural features such as structural type, construction year, window layout, and material composition. These models rely on deep learning tools, including Vision Transformers and CNNS. (Gonzalez et al., 2020; Ogawa et al., 2023; Szcześniak et al., 2022; S. Wang & Han, 2024). These methods have been applied across a variety of use cases. Ogawa et al. (2023) proposed a pipeline that classifies buildings' structural type and construction year by linking Street View images to cadastral footprints, filling gaps in regions where official records are incomplete. Gonzalez et al. (2020) extended this approach by estimating not only morphological characteristics such as building height and floor count, but also classifying structural typologies based on facade features, supporting broader analysis of urban form and building stock. More recently, Raghu et al. (2023) and S. Wang and Han (2024) have built on this work by identifying detailed facade characteristics such as transparency, openness, material composition, and signs of visual



Figure 2.17: Model attention visualized via CAMs (Raghu et al., 2023).

decay. Their methods rely on semantic segmentation and deep learning, as illustrated in Figure 2.17. These visual attributes help assess urban quality and evaluate thermal performance at the material level. Apart from their analysis, the researchers also point toward new possibilities for linking street-level imagery with sustainability goals and circular economy strategies, especially those focused on resource efficiency and the reuse of architectural elements.

A second dimension centers on *environmental simulation*. Instead of relying on generalized assumptions, researchers have begun to use visual data, particularly window-to-wall ratios (WWRs) and thermal imaging, to improve the accuracy of daylighting and energy performance models. For instance, Szcześniak et al. (2022) developed an automated method to extract WWRs from Google Street View imagery and validated it on more than 1,000 buildings in Manhattan in Figure 2.18. When used in simulations for Chicago, these actual WWR values increased predicted daylit areas by 9.5 percent and changed which buildings were prioritized for retrofitting. Building on this line of work, Klimenka et al. (2025) introduced a neural network that estimates facade surface temperatures using only RGB imagery and basic weather data. By avoiding the need for detailed geometric modeling, their approach makes thermal assessments faster and more scalable. Together, these studies show how street-level imagery can support more grounded and responsive environmental simulations at the urban scale.

A third domain involves the quantification of pedestrian experience and perceptual qualities. One of the earliest applications of this imagery has involved the modeling of visual perception. Publicly annotated datasets such as MIT's Place Pulse have been used to train

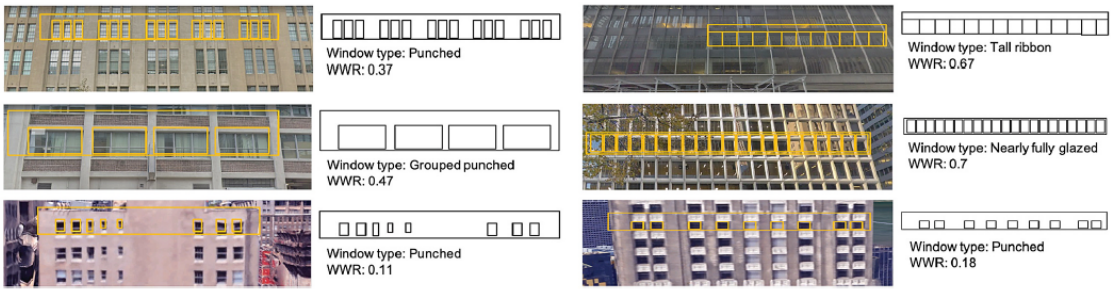


Figure 2.18: Detecting window-to-wall ratios for urban-level analysis



Figure 2.19: MIT Place Pulse dataset and the scores given in six dimensions (F. Zhang et al., 2018)

machine learning models to assign perceptual scores such as safety, liveliness, or walkability to street scenes in Figure 2.19 (F. Zhang et al., 2018). These perceptual scores have been mapped across entire cities, producing spatial representations that reflect emotional or aesthetic responses rather than physical infrastructure. Unlike conventional GIS layers, these representations introduce an affective dimension into urban analysis. Research such as (Fan et al., 2023) has shown that this kind of modeling is already informing decisions related to streetscape design and public investment. Deep learning segmentation applied to street imagery allows for the extraction of urban features relevant to walkability, including greenery, openness, and clutter. These perceptual elements have been operationalized into indices—such as the Visual walkability Index proposed by Zhou et al. (2019)—which correlate not only with spatial layout but also with socio-economic disparities. This framework has also been extended to include tasks such as vegetation indexing, land value prediction, and functional classification of buildings based on facade features (Li et al., 2022; Tang et al., 2021).

In addition to studies that address the perceptual, environmental, and experiential dimensions of urban space, a complementary line of research has concentrated on the detection of physical transformations within the built environment. This area is variously referred to in the literature as *urban transformation*, *physical transformation*, or more broadly, *change detection*. Within this domain, three interrelated methodological paradigms have emerged: *core-based models, semantic or class-based models, and hybrid approaches that combine elements of both* in Table 2.2.

Score-based approaches quantify urban change by assigning scalar or ordinal values derived from image features. These methods typically utilize regression or ranking models trained on human-annotated perception scores, which are safety, wealth, and cleanliness, or on self-supervised feature distances. A representative dataset for this purpose is Place Pulse. For instance, in the work of Z. Wang et al. (2024), the Place Pulse dataset was employed to evaluate perceptual changes in street view imagery (Huang et al., 2024). Their approach involved using deep convolutional neural networks to predict six perceptual attributes: safety, liveliness, beauty, wealth, depression (reframed as cheerfulness), and boredom (reframed as interestingness). These approaches are particularly effective in modeling perceptual shifts (e.g., gentrification) or aesthetic improvements but lack granularity in terms of what changed. They also often abstract away from the material composition of change (e.g., sign replacement vs. facade extension) by collapsing transformation into a single scalar value.

In contrast, semantic or class-based models attempt to explicitly categorize the type of transformation by assigning each detected change to a predefined class or label. These labels may be visual-functional (e.g., ”residential”, ”commercial”), or typological (e.g., ”signage added”, ”greenery removed”). Such methods typically rely on segmentation networks, object detectors, or captioning-based vision-language models, and often yield pixel-level class outputs. For instance, Y. Zhang et al. (2023) introduced a two-layer spatially dependent Graph Neural Network (GNN) model to classify urban functions from time-series street view imagery. Their model combines a semantic “knowledge” layer, built from vision-language descriptions, with a topological layer that incorporates spatial dependencies among street segments. This structure allows for the classification of each street segment into one of ten functional categories (e.g.,



Figure 2.20: Urban renewal probability across the street network predicted by the GNN model. Road segments are color-coded by renewal intensity: green (¡0.4), yellow (0.4–0.6), orange (0.6–0.8), red (¿0.8). Data source: OpenStreetMap (Y. Zhang et al., 2023).

commercial, residential, education) and enables the detection of function shifts over time in Figure **??**.

For instance, deconvolutional networks, as demonstrated by Alcantarilla et al. (2018), can be used to detect precise pixel-level changes by comparing deep feature reconstructions extracted from time-series street view imagery. Their model identifies altered regions by analyzing discrepancies in the activation maps of temporally aligned image pairs, enabling the localization of structural transformations such as building additions or facade removals. This approach produces fine-grained and interpretable outputs, making it particularly effective for applications that demand visual accuracy, including urban redevelopment monitoring and heritage conservation assessments.

One study by C. Liu and Song (2024) introduces a hybrid framework that integrates both visual similarity modeling and semantic modeling to detect urban redevelopment from time-series street view imagery. Their model operates through three parallel components: the Deep Pyramid Similarity Model (DPSM), Patterns and Processes of Urban Redevelopment (PPUR),

Authors (Year)	City/Region Studied	Imagery Source	Detection Type	Model/Method
Sakurada & Okatani (2015)	Japan	Custom street image pairs	Score-based (feature distance)	CNN + superpixel segmentation
Xie et al. (2016)	Not specified	Google Street View	Semantic	Deconvolutional Network (DeconvNet)
Wang, Ito & Biljecki (2024)	New York City, Singapore	Google Street View	Score-based (perception scores)	ResNet101 (DCNN)
Naik et al. (2017)	5 US cities	Google Street View (2007, 2014)	Score-based (difference metric)	SVM on semantic segmentation features
Qi et al. (2023)	Memphis, TN	GSV + aerial + LiDAR/SAR	Hybrid (semantic + geometric)	CNN pipeline
Zhang et al. (2023)	Global (China, mostly)	Tencent Street View	Semantic	Vision-Language model + GNN
Liu et al. (2025)	New York City, Memphis	Google Street View	Score-based (texture similarity)	Dual VGG16 + SVM classifier
Huang et al. (2024)	Seattle	Google Street View	Semantic + Temporal	Transformer-based (Street-BYOL, DINOv2)
Thackway et al. (2023b)	Sydney	Google Street View	Semantic + Socio-environmental	Deep learning + social indicators

Table 2.2: Overview of urban transformation detection studies using street-level imagery.

and the Markovian Support Vector Machine (MSVM). DPSM extracts multi-scale visual similarity features from segmented building pixels using a Siamese VGG16 architecture, emphasizing structural changes while suppressing transient elements like vehicles or trees. PPUR, on the other hand, evaluates the spatial and temporal plausibility of these detected changes by examining whether similar redevelopment patterns occur across nearby areas and over coherent time spans. It incorporates principles of geographic proximity and penalizes unrealistic or isolated transformation signals. The MSVM component then fuses the outputs of both DPSM and PPUR into a decision-making model, using a composite kernel that combines visual similarity with contextual consistency. This layered architecture allows the system to benefit from the smoothness and scalability of score-based visual metrics while grounding the classification in semantically and socially plausible urban processes. Unlike models that label changes into predefined semantic categories such as structure, facade, or function, Liu and Song's approach focuses on distinguishing plausible redevelopment events through the synergy of perception-driven similarity and spatial-process reasoning.

Despite methodological advances, current approaches to urban transformation detection often remain limited in scope and depth. Many models reduce change to scalar metrics or categorical shifts without capturing the underlying socio-spatial processes or symbolic meanings embedded in the built environment. The absence of localized *semantic reasoning*, especially

in distinguishing between types of physical change, constrains interpretability and operational use. Furthermore, most frameworks treat change as a binary or continuous signal, overlooking how certain transformations may be subtle and layered.

However, these operational uses have not been free of constraint. Limitations related to image quality, seasonal variation, and occlusion continue to restrict applicability. Geographic and socio-political biases in coverage are also present, as image collection has been concentrated in specific regions and urban centers. In addition, questions of privacy, surveillance, and algorithmic accountability remain unresolved. As visual processing becomes more integral to governance, the implications of these biases and omissions must be addressed.

Chapter 3

Karaköy Through Time: Empire, Republic and Redevelopment

3.1 Karaköy as a Site of Transformation

Karaköy was selected as the empirical focus of this study due to its layered urban morphology, continuous historical occupation, and distinct role in Istanbul's broader spatial-political economy. As a site that has accommodated diverse infrastructural and functional regimes—from Byzantine maritime facilities and Genoese fortifications to Ottoman trade networks and early Republican financial institutions—Karaköy constitutes a rare case of uninterrupted urban development. This layering enables diachronic analysis while simultaneously making it possible to observe overlapping spatial transformations (Erbil & Erbil, 2001; Okyay & Binan, 2017).

Importantly, Karaköy has remained a visually and experientially accessible site throughout these changes. It is a location historically subject to direct observation by everyday users. Since the 1980s in particular, the area has undergone gradual processes of deindustrialization, capital inflow, and aesthetic rebranding, positioning it as one of the early spatial expressions of Istanbul's engagement with market-oriented urban restructuring. In this sense, Karaköy reflects the long-term embedding of capitalist urban logics within the city's central zones (Karaman, 2008).

Karaköy also satisfies the methodological demand for a rapidly transforming urban environment. Between 2000 and 2025, the area has experienced intensive physical and symbolic reconfiguration, including large-scale privatization of the waterfront and changes in land use. These dynamics make it suitable for detecting high-frequency transformation patterns (Tepeli

& Ocakçı, 2017).

From a data perspective, Karaköy is one of the few districts in Istanbul with dense, temporally consistent Google Street View coverage, enabling the construction of visual time series suitable for computational image-based analysis. Furthermore, its spatial compactness, representational density, and socio-political centrality render Karaköy an empirically traceable site and a contextually rich urban case through which broader theoretical questions on morphology, capital, and representation can be operationalized.

3.2 Late Ottoman Period

Over the centuries, Karaköy has not been mentioned separately from Galata; rather, historical sources typically refer to it as the port or waterfront area of Galata. In this study, therefore, the broader historical designation *"Galata"* will consistently be employed to encompass both the hillside settlement and its waterfront extension, currently known as *Karaköy*. During the Genoese occupation, this waterfront was bounded by substantial fortification walls that encircled the entire colony, reinforcing Galata's territorial integrity as a semi-autonomous maritime outpost (Küçük & Mazlum, 2017).

Galata, initially known as Sykae—meaning "fig garden"—during the Byzantine era, emerged as a strategically positioned settlement directly across from Constantinople. Its prominence derived from its location along the northern shore of the Golden Horn, a site that allowed maritime access, defensive advantage, and urban expansion (Arseven & Yelkenci, 1989). By the reign of Emperor Theodosius in the late 4th century, the region—referred to as Regio Sycaena—had developed into an enclosed neighborhood complete with a church, public forum, baths, shipyards, theater, bakery, mill, covered portico, and an estimated 431 houses (Güleçyan, 2017). This illustrates an urban fabric within the city's 13th administrative district. Over subsequent centuries, Sykae evolved into Sycia and eventually medieval Galata, defined by its crescent-shaped topography that extended from the Galata ridge down to the harbor basin, locked between the Bosphorus and the Golden Horn (Okyay & Binan, 2017)[1].

[1]The topography of Galata, descending from a northern ridge toward the harbor level in a concave form, played

Throughout the Byzantine period, Galata functioned primarily as Constantinople's commercial maritime quarter. Byzantine Galata functioned as a vibrant waterfront area, shaped by the presence of workshops, piers, taverns, and custom houses—features that underlined its role as a logistical and mercantile zone (Okyay & Binan, 2017).[2] Genoese and Pisan merchants who operated along the Black Sea routes frequently used Galata as a trading station and storage point. The Byzantine state participated in this network by offering privileges to foreign traders and collecting customs revenues from places such as Hyperpera (Müller-Wiener, 2001). Starting in the 11th century, Galata began to grow as a commercial hub, drawing merchants from Italian city-states like Genoa, Venice, and Amalfi. These traders were offered significant privileges in return for helping to move goods across the Mediterranean. While the Genoese population was predominantly Catholic, the area was also inhabited by Orthodox Christians, Armenians, and Jews. Churches, cemeteries, and similar religious institutions served as physical expressions of this diversity, shaping the architectural character of Galata. The presence of both Latin Catholic and Eastern Orthodox buildings highlights the coexistence of distinct cultural and religious traditions within the district (Müller-Wiener, 2001). [3] [4]

After the Fourth Crusade in 1204, Venetian influence over Galata grew as the Latin Empire took shape. But this shifted in 1261, when the Byzantines reclaimed Constantinople and granted the Genoese exclusive control over Galata. From that point on, the district began to change: the Genoese fortified it, placed it under the authority of their own appointed Podestà, and gradually turned it into a city within a city.[5] Galata's layout resembled Latin city models and was based on a grid plan. The Genoese established their own legal system, commercial regulations, and defensive infrastructure within the colony.[6] The construction of the Galata Tower

a crucial role in shaping the district's defensive walls, maritime infrastructure and intra-urban circulation patterns. This geographical contour directly influenced neighborhood organization and architectural orientation across historical periods (Okyay & Binan, 2017).

[2]Okyay bases this description on Müller-Wiener's account of Galata in the Byzantine period, particularly as outlined in Bildlexikon zur Topographie Istanbuls (2003, pp. 137–138), where the area is depicted as a lively port district defined by its commercial and infrastructural elements.

[3]Arseven refers to Belin's 1894 account in his discussion of Galata's early trade dynamics and privileges granted to Latin merchant communities. See Arseven (1989, p. 55).

[4]The multicultural structure can be traced in the varying architectural styles of sacred buildings, reflecting the coexistence of different faith communities. These religious forms contributed to Galata's diverse identity and formed the basis for the emergence of non-Muslim neighborhoods in the Ottoman era (Müller-Wiener, 2001).

[5]The Podestà was the official representative appointed by the Republic of Genoa to administer its colonies abroad.

[6]The construction of the fortification walls played a central role in defining Galata's territorial boundaries.

in 1348 stood as a clear marker of their rising power, both physical and economic. By the late medieval period, Galata was enclosed by stone fortifications, with controlled access provided through gates such as Mumhane and Kireç Gates (Kafesçioğlu, 2016). Galata evolved into a semi-autonomous district whose significance was defined by its intermediary position between the European capital and Eastern markets. However, this autonomy was shaped by shifts in imperial power and ultimately dissolved following the Ottoman conquest of Constantinople in 1453, which reconfigured the geopolitical order of the region.

After the Ottoman conquest of Constantinople in 1453, Galata surrendered to Sultan Mehmed II without resistance. Although the Genoese had formally declared neutrality, their cooperation with the Byzantines during the siege was noted by the Ottomans. As a consequence, parts of Galata's fortifications were dismantled and the upper sections of the Galata Tower were damaged in what appeared to be a symbolic assertion of Ottoman authority (Arseven & Yelkenci, 1989). The district lost its political autonomy and was redefined within the Ottoman system as a non-Muslim communal entity, or cemaat-i diniyye. Yet Galata's location and commercial infrastructure continued to matter. Acknowledging Galata's continued strategic and economic significance, the Ottoman administration reinstated many of the commercial privileges previously granted under Genoese and Byzantine rule. As a result, European presence in the district remained substantial. The development demonstrates the persistence of Galata's function as a key node of imperial trade and cross-cultural diplomacy under Ottoman governance (Kafesçioğlu, 2016). Although Galata maintained its commercial function under Ottoman rule, the spatial development of both the district and the wider city remains difficult to trace due to the limited availability of systematic cartographic documentation from the early modern period.

The urban morphology of the 17th and 18th centuries, Istanbul, has not been systematically documented in terms of social transformation or architectural character. Historical cartographic analyses rely on maps produced after 1858, leaving the urban fabric of the pre-nineteenth-century period without visual support. The earliest maps of Istanbul drawn by the

These walls provided military protection and contributed to the spatial and symbolic distinction of Pera from the rest of Istanbul. Gates and bastions marked a formal boundary that emphasized Galata's semi-autonomous status and reinforced its unique urban identity during the late Byzantine and early Genoese periods.

Ottomans were miniatures, often referred to as pictorial maps. A scaled cartographic representation encompassing this area does not appear until the nineteenth century. One of the most prominent early depictions is found in Beyân-ı Menâzil-i Sefer-i Irâkeyn by Matrakçı Nasûh (d. ca. 1564), where a miniature illustration of Istanbul provides a topographical portrayal characterized by distinctive features of the city in Figure 3.1 (Dağdelen, 2021). [7]



Figure 3.1: Miniature depiction of Istanbul from Matrakçı Nasûh's Beyân-ı Menâzil-i Sefer-i Irâkeyn (Dağdelen, 2021).

Although detailed maps and architectural records from the 17th and 18th centuries are limited, this period witnessed notable changes in the spatial configuration of Galata, particularly in its lower sections corresponding to present-day Karaköy. While the area remained encased within its Galata fortifications, these walls gradually lost their military function as the boundaries between defense and urban productivity began to dissolve. By the 16th century, the gates of Galata, such as Mumhane, Kireç, and Eğri Kapı, no longer served just as entry points. The fortifications had become focal nodes of artisanal production. Around these thresholds,

[7] This depiction includes both "nefs-i İstanbul" (the core of the city) and its associated judicial districts—Galata, Eyüp, and Üsküdar— referred to as the Bilâd-ı Selâse.

candlemakers rendered animal fats, tanners processed hides, and powder-makers supplied imperial armaments. These activities transformed the outer belt of the walled city into a porous frontier of industry and exchange (Okyay & Binan, 2017).

By the 18th century, the spatial logic of permeability had become more pronounced. Formerly defensive walls, once associated with boundaries and control, had begun to take on new functions within the urban economy. Residential structures were allowed to be built on top of or alongside them, often in exchange for fees. Fortifications were progressively absorbed into the city's morphological structure in response to patterns of urban expansion and spatial reorganization. At the same time, remnants of Byzantine infrastructure were recontextualized within the framework of Ottoman urban practices and spatial logic (Küçük & Mazlum, 2017). [8]

In the 18th century, small-scale civic and commercial buildings became more common across the city. The modest hans built near the harbor in 1732 articulated the semi-public nature of the area's infrastructure and reflected imperial attention to the organization of everyday life (Okyay & Binan, 2017). This gradual densification, preceding formal land reforms, set the stage for the major spatial transformations of the 19th century. In the early phases of Ottoman modernization, Karaköy was no longer a fortified edge of the city, but had become a space of transition shaped by changing forms of imperial control and ways of understanding urban space.

By the late nineteenth century, the transformation of Karaköy was shaped by large port construction projects and land reclamation along the coastline. The area's previously irregular and organically formed shoreline was replaced by a straight, engineered edge. New office buildings and bonded warehouses were built along this redesigned waterfront. (Küçük & Mazlum, 2017). This transformation was not limited to physical restructuring; it also signaled a functional and economic redefinition of the area, aligning Karaköy more directly with international trade networks. The process unfolded within the broader context of late Ottoman modernization, shaped by a series of interrelated forces. Recurrent urban fires created opportunities for spatial reorganization, military conflicts generated both destruction and new infrastructure

[8] One example of such a transformed structure is the *Kastelion ton Galatou*, a subterranean space of uncertain origin that was reinterpreted in the Ottoman period as the *Mahzen-i Sultani* and later repurposed as *Yeraltı Camii*. Remarkably, the structure has survived and continues to function as a mosque today (Okyay & Binan, 2017).



Figure 3.2: Goad's 1905 fire map shows building materials and functions west of Galata Bridge, near port structures (No:26) (Goad, 1905).

demands, and the increasing presence of foreign capital facilitated the implementation of modern urban interventions. These changes are documented in three key cartographic sources: *G. d'Ostoya's 1858 map, Charles Edouard Goad's fire-insurance atlas from 1905* (Bahçekapılı, 2020; Küçük & Mazlum, 2017). [9]

Throughout the 19th century, Karaköy experienced a series of fires, particularly in the years 1840, 1852, 1857, 1874, and 1957. These disasters destroyed most of the timber houses and accelerated the transition to fire-resistant masonry construction. In Goad's maps, this material change is represented through the use of pink tones, often indicating masonry buildings in Figure 3.2. These features were introduced to enhance fire safety and also embodied a change in architectural expression shaped by the priorities of finance and insurance sectors (Küçük & Mazlum, 2017). At the street level, traditional timber houses that had once combined domestic and commercial uses were replaced by purpose-built hans. These multi-story buildings accommodated banks, insurance agencies, and architectural offices on their upper levels (Kafesçioğlu, 2016).

[9]D'Ostoya's survey underpinned some of the earliest municipal regulations. Goad's detailed renderings were used by British and French insurers to assess risk, building materials, and susceptibility to fire.



Figure 3.3: Timber Vitalis Han ve Ottoman Bank Building on Voyvoda Street (SALT, 2021).

As Karaköy modernized, some of its most defining physical and symbolic boundaries were removed. Sections of the old Galata Walls, which had once marked the edge of the medieval settlement, began to be dismantled in the 1860s. This made it possible to open new boulevards such as Kemeraltı and Necatibey Caddesi, connecting the modernized port area with the expanding financial center around Voyvoda Street (Kafesçioğlu, 2016).

From the 1860s until the end of the century, its transformation was shaped by the arrival of financial, technological, and legal firms in Voyvoda Street. Banks, insurance firms, legal offices, architectural firms, mining companies, railway agencies, and technology importers are concentrated along this street. (Eldem, 2000; Kafesçioğlu, 2016).[10]Although only around 300 meters in length, the street hosted forty buildings whose neoclassical facades and rusticated stone bases visually embodied the authority and prestige of finance in Figure 3.3.[11]

In the late nineteenth and early twentieth centuries, the mixed-use character of Karaköy, where residential, artisanal, and small-scale commercial activities had long coexisted, began to change. The area underwent a process of functional specialization and spatial reorganization. By 1905, masonry buildings had largely replaced timber structures, reflecting both a material and architectural transformation. A new building type, the han, became widespread. These

[10]Voyvoda Street, later officially renamed as Bankalar Street.

[11]Voyvoda Street once housed several historically significant buildings, some of which no longer exist due to later urban interventions such as the Menderes-era demolitions. Notable examples include Nur Han (demolished during the Menderes redevelopment), the Vakıflar Bankası building, Jeneral Han, the Sümerbank building, Voyvoda Han, the Akbank building, Hezeran Han, and the iconic Kamondo Stairs (Eldem, 2000).

multi-storey commercial buildings followed a clear vertical organization: ground floors were used for retail or shipping, upper floors for offices, basements for storage, and in some instances, chapels or clubs were situated on the rooftop (Kafesçioğlu, 2016). This architectural ordering introduced new spatial hierarchies aligned with bureaucratic and commercial rationality. As a result, older family-operated enterprises were displaced toward peripheral zones such as Tophane and Kasımpaşa, while a new class of professionals, including clerks and typists, began to inhabit these increasingly institutionalized environments.

This typological and socio-economic transformation was supported by infrastructural developments intended to regulate urban mobility and communication. The Tünel, an underground funicular opened in 1875 to connect Karaköy with the Pera ridgeline, played an important role in improving vertical circulation within the city (Erkökoğlu, 2002). In addition, the extension of tramlines, the installation of sewer systems, and the erection of telegraph poles contributed to the formation of a more integrated and regulated circulation network (Küçük & Mazlum, 2017). Nevertheless, despite these modernizing processes and the tendency toward spatial homogenization, Karaköy preserved its long-established ethno-religious diversity throughout this period. Greek Orthodox, Armenian, Jewish, Levantine, and Muslim populations continued to reside in shared urban blocks. Their places of worship and communal institutions were systematically recorded in the detailed cartographic surveys produced by Goad and, subsequently, by Nirven.[12]

Karaköy changed many times between the Byzantine period and the late Ottoman era. Fortification walls were removed, timber houses were replaced by masonry hans and new systems such as the Tünel, tramlines and sewers were installed. These projects gave the district a modern street grid and new building types. Despite these transformations, the district maintained its heterogeneous social fabric, with multiple religious and ethnic groups sharing the same region. Karaköy illustrates that urban change did not begin with the Republican era but had earlier roots shaped by shifts in infrastructure and spatial organization. Its layered urban

[12]The Nirven maps refer to a set of detailed urban cartographic surveys produced by Suat Nirven between 1948 and 1949 as a continuation of the earlier Pervititch insurance maps. These maps documented functions and structural types in central districts of Istanbul, including Karaköy, and are considered key resources for understanding 20th-century urban morphology (Küçük & Mazlum, 2017).

fabric and visible transformations provide a clear basis for studying long-term patterns of morphological development.

3.3 Early Republican

The early Republican period marked a significant phase of state-led urban intervention in Karaköy. Despite the relocation of the capital to Ankara, Karaköy maintained its importance in Istanbul's commercial activities due to its strategic position along the waterfront. In an effort to break away from the Ottoman past, the new Republic aimed to establish a modern and well-organized urban environment shaped by Western planning principles. One of the first major steps was the 1933 international planning competition, which attracted figures such as Henri Prost, Jacques-Henri Lambert, Donat-Alfred Agache, and Hermann Ehlgötz. Although Ehlgötz's scheme won, it was shelved; instead, Prost was hired in 1936 to draft Istanbul's first comprehensive master plan. His 1937 plan emphasized wide boulevards, expansive green areas, and a coordinated approach to urban aesthetics. Within this vision, Karaköy was reimagined as a modern and orderly district.

Prost's vision was layered over an already active and functionally complex region. The district still accommodated a dense concentration of banks, stock exchanges, and post offices, accompanied by the expansion of office buildings and insurance institutions. Han-type commercial structures housed specialized professional services, including banking, law, architecture, and engineering, illustrating the area's diversified economic profile. Alongside these large-scale institutions, small-scale retail and service-oriented businesses—such as barbers and tailors—contributed to the daily functioning of the neighborhood. Furthermore, Karaköy retained its significance as a port zone through ongoing maritime and logistical operations, particularly in the Salıpazarı vicinity (Kafesçioğlu, 2016; Küçük & Mazlum, 2017) . The Galata Quay was designated to retain its function as a passenger port, while new streets and public squares were planned to better integrate Beyoğlu, Karaköy, and Tophane into the city's transportation network in Figure 3.4 (Bilsel, 2011; Erkökoğlu, 2002).

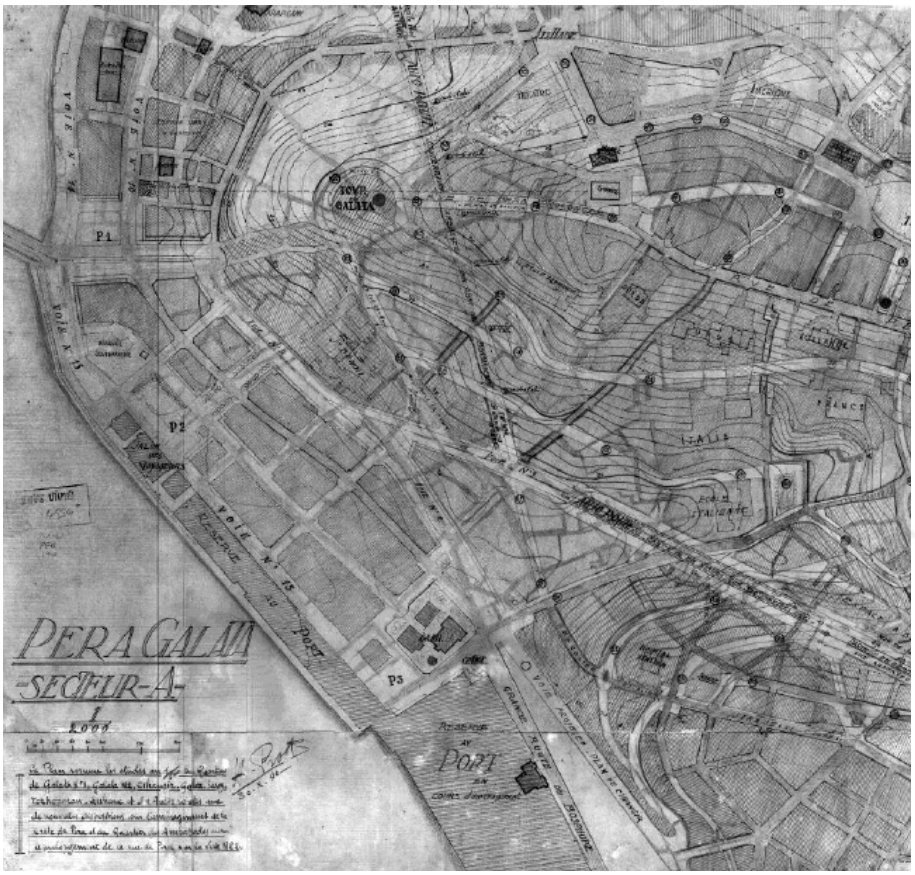


Figure 3.4: An illustration of Galata's transformation as envisioned by Prost, taken from his 1/2000 scale Galata-Pera Plan.

During the 1950s, Prime Minister Adnan Menderes initiated a series of large-scale urban renewal operations known as the *"Menderes demolitions"*. These interventions were inspired by Prost's earlier plans but lacked their design sensitivity and coherence (Bilsel, 2011). Major projects such as the 30-metre-wide Vatan and Millet boulevards followed the alignments originally drawn by Prost. However, their implementation involved the displacement of many residents and significant destruction of historic buildings. One of the most prominent actions was the widening of Kemeraltı Street to approximately 30 meters, which required the demolition of 92 buildings and the expropriation of around 300 properties (Küçük & Mazlum, 2017). Approximately 92 buildings were either partially dismantled or fully demolished. These included Kabataş Hamamı, the shops surrounding Kılıç Ali Paşa Mosque, and many 19th-century facades (Okyay & Binan, 2017). Earlier road construction in the same decade had already led to the removal of Christos Rum Church. In addition, three nearby Orthodox churches—Aziz Yahya, Merkez Meryem Ana, and Aziz Nikola—were transferred to the newly established Turkish Orthodox Patriarchate after 1922, which altered the religious landscape of the district (Küçük & Mazlum, 2017). The destruction extended to several commercial buildings and hans that once defined Karaköy's architectural identity. Structures such as Havyar Han, Borsa Han, Noradunkyan Han, Kevork Bey Hanı, Manukoğlu Hanı, Mehmet Ali Paşa Hanı, and the Aziziye Police Station were among those lost during this period (Bilsel, 2011; Eldem, 2000). In 1951, Prost resigned from his position, citing growing speculative pressures and the disregard for his planning principles. Following his departure, urban planning in the city moved away from a coordinated approach and gave way to fragmented interventions that often caused lasting damage to the historical fabric (Bilsel, 2011).

In the 1960s, the district lost its central role as Istanbul's primary port due to the increasing importance of Haydarpaşa Port. As a result, Karaköy's commercial activity declined. However, after the dissolution of the Soviet Union in 1991, the area experienced a temporary revival through the so-called "luggage trade," as visitors from former Soviet republics engaged in informal cross-border commerce (Erbil & Erbil, 2001).

In conclusion, Karaköy's transformation has been shaped by changes in political priorities and economic conditions. The early Republican period played an important role in setting



Figure 3.5: A 1958 image capturing a demolition activity near Karaköy Square, extending toward Bankalar Street. Noradunkyan Han was demolished. (Eldem, 2000)

the stage for this transformation. Efforts to modernize the district and move away from the Ottoman past led to the adoption of Western planning ideas, especially through Henri Prost's 1937 master plan. Although this plan was only partially realized, it introduced a new way of thinking about urban space and created a framework that influenced later developments. In the 1950s, this foundation made it easier for the Menderes government to carry out large-scale interventions, although often with less concern for historical preservation. In the following decades, Karaköy continued to change as port activity declined, new trade patterns emerged, and global economic pressures increased. These changes have kept Karaköy at the center of ongoing tensions between development and heritage conservation.

3.4 2000s and Beyond

In the 2000s, Karaköy entered a new period of change shaped by global economic forces and neoliberal urban policies. The district, once defined by its port functions and financial industry, became a target for large-scale redevelopment. Its central location and waterfront

made it attractive for tourism and commercial activity. Starting in the 1980s and gaining speed in the early twenty-first century, planning approaches began to focus more on private capital and consumption rather than public infrastructure and everyday use. As a result, Karaköy became an example of how global city strategies can reshape urban space, often creating tension between economic development and the protection of local identity and historical continuity (Erbil & Erbil, 2001; Tepeli & Ocakçı, 2017).

However, the roots of this transformation did not emerge suddenly in the 2000s. The roots of this transformation can be traced back to the 1980s, when economic and political changes on both national and global scales began to reshape Istanbul's urban priorities. The widespread adoption of neoliberal policies led to a strategic shift away from state-led industrialization towards market-driven development (Aksu & Tutar, 2021). For Istanbul, a historic industrial and maritime hub, this meant a period of painful deindustrialization. The traditional economic engines of Karaköy, including its bustling port, ship-chandlers, artisan workshops housed in hans, and small-scale manufacturing, began to decline. As maritime trade modernized and moved to larger, more peripheral facilities, the district's prime waterfront real estate, including its historic Ottoman-era warehouses and early Republican buildings, was increasingly framed by developers and policymakers as economically obsolete. In urban planning discourse, these historically significant spaces were increasingly described as *"underutilized"* and *"ripe for redevelopment"*. Such terminology tended to overlook their ongoing, though reduced, economic and social functions for the local population (Erbil & Erbil, 2001).

This economic change was constructed by influential theoretical perspectives through which the transformation of *physical space into a consumable product* can be understood. A foundational argument was articulated by Lefebvre (2012), who asserted that space should not be seen as a neutral backdrop for social life. Instead, it is shaped and reorganized by the dominant economic systems. In capitalist contexts, space is transformed through processes that turn it into a commodity (Aksu & Tutar, 2021; Lefebvre, 2012).

Yet this transformation extends beyond the consumption of space. The district itself is not only

used for consumption but is reconstituted as an *object of consumption* (Bahçekapılı, 2020; Foucault, 2013). According to Foucault (2013), objects—including places—are not assumed to possess inherent or fixed identities; rather, they are brought into being through historically contingent discourses. Through discourse, a place is defined and problematized within a framework that renders it intelligible and governable.

Likewise, David Harvey's analysis of neoliberalism highlights its manifestation through urban restructuring, where cities are re-engineered to facilitate capital accumulation. In this context, Karaköy's historic identity became both a liability and an asset. Its industrial past made it a target for "renewal," while its rich architectural heritage and waterfront location made it an invaluable resource for a new kind of economy based not on production, but on consumption, tourism, and spectacle (Aksu & Tutar, 2021). Building on these socio-economic shifts, the announcement of the Galataport project in the early 2000s marked a decisive turn in Karaköy's urban trajectory. If the transformations of the 1980s and 1990s laid the groundwork, the announcement of the Galataport project in the early 2000s marked an important turning point in Karaköy's transition into a new urban configuration. Framed as a solution to the district's so-called economic stagnation, the project aimed to transform the historic Salıpazarı port into a global hub for tourism and high-end commerce, spanning approximately 1.2 kilometers of public coastline (Karaman, 2008). Although the area had already been designated a Tourism Zone in 1994 and zoning plans were approved in 2004, the project faced substantial opposition. The project was met with objections due to concerns about excessive building density, the scarcity of accessible public space, the creation of a barrier between the city and the waterfront, and the potential threat to the district's historical fabric. In response, professional chambers—including those of architects, urban planners, and civil engineers—filed lawsuits that led to a court-ordered suspension of the project in 2006 (Erbil & Erbil, 2001; Küçük & Mazlum, 2017).

The project re-emerged in the aftermath of legal and administrative restructuring. Following amendments to the Coastal Law in 2005 and the area's inclusion in the national privatization portfolio, the Salıpazarı port was transferred to the authority of the Privatization Administration (ÖİB). A new Environmental Impact Assessment (EIA) was approved in 2015, paving the way for the clearance of the site and the demolition of several historical port structures, including



Figure 3.6: Satellite image of the Galataport project area along the Karaköy coastline, outlined in red (Şehir Planlama Müdürlüğü, n.d.).

customs warehouses and storage buildings (Bahçekapılı, 2020; Tepeli & Ocakçı, 2017).

Originally introduced as the *"Salıpazarı Cruise Port"* in 2002, the project gathered momentum in the 2010s. Between 2016 and 2018, intensive construction work reshaped the district's built environment in Figure 3.6. Some structures were preserved and refunctionalized, while others were demolished or replaced with new commercial spaces. Field studies document that at least 27 deteriorated buildings underwent redevelopment interventions during this period (Sarı et al., 2024). Galataport was officially completed and opened to the public in 2021 (Bahçekapılı, 2020).

The redevelopment introduced profound spatial transformations. The interior and roof of the Karaköy Passenger Terminal were removed in Figure 3.7, while its historical facade was preserved and incorporated into the newly constructed Peninsula Hotel in Figure 3.8. The Parcel Post Office building was largely dismantled, retaining only its front facade. Several customs buildings and warehouses along the waterfront were demolished to make way for new museums, public walkways, and retail establishments. A large-scale underground cruise terminal, the first of its kind in Istanbul, was excavated to relocate port functions below ground in order to maximize surface-level space for leisure and consumption. While certain heritage buildings such as



Figure 3.7: The former Karaköy Passenger Terminal before its demolition as part of the Galataport redevelopment, representing Istanbul's mid-20th century port infrastructure (Şehir Planlama Müdürlüğü, n.d.).

Çinili Han and Merkez Han were restored and adapted for new uses, the broader transformation led to the erasure of much of Karaköy's original maritime infrastructure and spatial memory. These interventions reignited debates surrounding the commodification of cultural heritage and the privatization of formerly public urban spaces (Şehir Planlama Müdürlüğü, n.d.).

Beyond the material interventions and architectural rearrangements, the redevelopment was also ideologically charged, shaped by a vision of urban competitiveness and place branding in the context of neoliberal globalization. The Galataport development plan envisioned an integrated zone combining cruise ship terminals, luxury hotels, high-end retail, fine-dining venues, and cultural institutions such as the relocated Istanbul Museum of Modern Art, with the stated aim of positioning Istanbul as a key destination in the global luxury economy by attracting affluent international visitors and enhancing the city's image and competitiveness (Karaman, 2008). Supporters of the project, primarily institutional and class-based actors, included those who stood to benefit from neoliberal urban redevelopment, particularly public–private coalitions composed of municipal planning bodies, private investors, and tourism-sector stakeholders. These groups emphasized expected benefits such as foreign investment, job creation,



Figure 3.8: The design proposal for the Peninsula Hotel, developed by the Hong Kong-based hospitality group (Şehir Planlama Müdürlüğü, n.d.).

infrastructure upgrades, and increased tourism revenues (Aksu & Tutar, 2021), framing Galataport as a symbol of Istanbul's aspiration to join the ranks of global cities that had revitalized their waterfronts for economic renewal. Yet beneath this promotional narrative, a more complex reality emerged, as the transformation introduced new economic opportunities and visual enhancements but also provoked concerns about displacement and the erosion of public space. The project became a focal point in ongoing debates about for whom the city is being remade and what is gained or lost when urban space is reshaped according to the logics of tourism and investment" (Aksu & Tutar, 2021; Karaman, 2008). However, the socio-spatial costs of this transformation were unevenly distributed, with pre-existing local communities and small-scale economic actors disproportionately affected. Their displacement constituted a significant yet often overlooked consequence of the project, revealing the asymmetries embedded within the celebrated narrative of urban renewal and global competitiveness.

The impact of these dynamics has become more visible after Galataport became operational. In the last five years, the proliferation of luxury cafés, restaurants, and hotels has made the area increasingly attractive to high-income groups. This shift has intensified socio-spatial

tensions between newly arrived affluent consumers and the low-income populations who previously sought shelter and informal livelihoods in the abandoned or derelict parts of the city center. As gentrification reconfigures formerly neglected urban cores, the resulting juxtaposition of wealth and poverty fosters new forms of urban conflict (Küçük & Mazlum, 2017).

The transformation of Karaköy extends beyond the physical boundaries of Galataport, instigating a broader process of gentrification that re-brands the entire district as a consumable experience. This process converges in the creation of what can be termed a ”design city”—an urban space meticulously curated to compete in the global marketplace for investment, creativity, and tourism. In this model, heritage and nostalgia are not simply preserved but are commodified and sold as experiences (Aksu & Tutar, 2021).

This rebranding process is most visible in the proliferation of stylized interiors, luxury restaurants, minimalist cafés, and boutique shops that have colonized the streets adjacent to the new development. These venues are not just material places; they are imagined realities, carefully crafted to appeal to a specific demographic with *distinct tastes and sensibilities*. Visual consumption becomes paramount. The urban experience is mediated through ambient aesthetics, photogenic signifiers, and curated imagery, perfectly suited for dissemination on social media platforms like Instagram. As venue owners have noted, generating engagement through ”likes” is a deliberate strategy to enhance visibility and drive consumption. These spaces function as sites of everyday social differentiation, reifying themes of status and luxury while fostering new, often exclusive, modes of socialization (Aksu & Tutar, 2021).

The aesthetic and spatial reconfigurations introduce stark spatial contradictions. The new privatized, secured, and gated spaces of Galataport stand in sharp contrast to the older, more organic urban fabric surrounding them. This deepens socioeconomic segregation, effectively creating invisible walls that exclude many local residents from the new developments (Bahçekapılı, 2020). The waterfront, once a porous and accessible public amenity, has been converted into an exclusive commodity, primarily accessible to wealthy locals and international tourists willing to pay for the privilege of consumption (Erbil & Erbil, 2001).

The lived experience of this new Karaköy is one of negotiation and contestation across

what one study calls a *spatial divide*. Walking-based analyses of the district reveal the powerful role of imaginary boundaries in structuring urban perception and social interaction. While Karaköy remains a space of encounter for diverse user groups—old-time shopkeepers, construction workers, tourists, and the new creative class—these encounters often serve to reproduce a sense of *otherness*. Social groups continually reframe one another as outsiders, leading to the repeated production of symbolic borders through everyday spatial practices (Bahçekapılı, 2020).

Karaköy serves as an example of how neoliberal urban transformation brings changes to the city's built environment, as well as to the social and symbolic dimensions of urban space. Triggered by changes in economic policy beginning in the 1980s, this transformation accelerated in the early 2000s and took concrete form with the implementation of the Galataport project. Through strategies of aestheticization, privatization, and urban rebranding, Karaköy was redefined as a curated site of global consumption and symbolic capital.. However, this process entailed significant contradictions. The selective preservation of architectural heritage, the commodification of collective memory, and the exclusionary mechanisms embedded in redevelopment practices have contributed to the emergence of new spatial fragmentations and social inequalities. While the project achieved its stated goals of attracting investment and enhancing the city's international image, it simultaneously produced symbolic and material boundaries that deepened class-based segregation and displaced long-established communities. As such, Karaköy exemplifies the structural tensions inherent in contemporary urban governance, where the pursuit of global competitiveness frequently undermines local continuity, inclusivity, and the everyday practices that sustain urban life.

Chapter 4

Data Acquisitions, Inferences and Results

4.1 Framework of Visual-Linguistic Interpretation

This thesis proposes a multi-layered methodological framework for detecting and interpreting facade-level urban transformations using Google Street View (GSV) imagery. Instead of viewing GSV images as neutral records, the study treats them as mediated representations shaped by both technical protocols and social contexts. A domain-specific ontology—comprising facade, structure, and function—guides the detection and classification of changes in the built environment. The methodology operates through three layers: (1) a computational layer that retrieves, aligns, and stitches historical panoramas into temporal grids for each building; (2) an AI implementation layer that uses a multimodal language model to generate structured outputs based on these images; and (3) an interpretive layer that situates detected transformations within their spatial and socio-morphological context with a meaning-making critique informed by Wood's conceptualization of mapping. Together, these layers enable a more nuanced reading of urban change that connects visual signals to broader socio-spatial processes. The data encompasses 3,542 GSV images and 781 collage frames, 843 GSV images and 113 collage frames for Galataport, and 130 GSV images and 26 collage frames for the coastline zone. [1]

Additionally, the inclusion of 19,272 Mapillary images and 284 historical or geocoded archival visuals—sourced from SALT Research and social media platforms—further expands

[1] The Appendix includes a visual archive interface developed for this study, in which all image sources—Google Street View, Mapillary, and Salt Research—are integrated and visualized simultaneously within a defined coordinate frame using a reduced-radius filter. This allows a comparative spatial reading of the dataset across sources at each transformation point.

the temporal depth and spatial granularity of the analysis. While these supplementary sources lack the systematic structure and temporal consistency of GSV imagery, they offer valuable contextual insight into platform-specific patterns of visibility and omission. Moreover, they contribute historical and morphological information that enhances the interpretation of spatial transformation, particularly in areas where formal records are limited or absent. Collectively, these heterogeneous visual materials support a more nuanced reading of urban change by linking localized visual signals to broader socio-spatial dynamics and representational structures.

To grasp urban transformation with street view imagery, it is important to pay attention to small visual changes, not as background scenery but as meaningful changes shaped by the area's social and historical context. The methodological workflow begins at the moment of data acquisition: every coordinate of metadata is queried through the Google Street View Metadata API, and the request returns not a single panorama but a chronologically ordered list of identifiers. The client, a custom Python script developed to interface with the Google Street View Metadata API, enforces a fixed heading, pitch, and field of view to ensure consistency across temporal image captures. Although this technical precaution appears unimportant, its consequence is significant: it prevents the inadvertent conflation of camera movement with urban change and thereby protects subsequent interpretation from artefactual noise. Each downloaded image is renamed according to a *strict schema that embeds latitude, longitude, heading, time stamp,p, and a unique hash into the filename*. Embedding context inside the file system allows downstream scripts to recover provenance without external look-up tables, and it transforms an otherwise directory of images into *a searchable, auditable archive*.

To overcome this constraint, the following three-step methodology has been developed:

1. **Retrieving Current GSV Metadata:** The process begins by using the Google Street View Metadata API to obtain URLs and pano IDs of the latest available imagery for selected locations.

2. **Extracting Historical Imagery:** Since Google Maps allows users to access historical Street View data, a custom approach is necessary. By constructing Google Maps URLs and utilizing web automation tools such as Selenium, historical pano IDs can be extracted

for further processing.

3. **Downloading and Standardizing Imagery:** Once historical pano IDs are obtained, images are retrieved at appropriate pitch and point-of-view (POV) settings, ensuring consistency across temporal datasets.

In the ontological framework for understanding urban transformation, an effort was made to capture both buildings and street infrastructure. To achieve this, both building-based and road-based datasets were integrated. Generally, researchers utilize open-source data such as OSM for locating building positions. However, in Karaköy, the building stock in the western part is highly irregular, making it difficult to accurately interpret building locations from top-down images. Therefore, instead of relying on OSM data, the municipality's CAD file was used. Both data sources were tested, and the accuracy of the CAD file was confirmed in Figure 4.1. Each building was then represented as a polygon in a shapefile, and its centroid was calculated. The nearest GSV point to this centroid was identified using API queries. A line was drawn from the GSV point to the centroid, and the viewing angle was determined based on Street View API's heading conventions (0° for North, 90° for West, 180° for South, and 270° for East). The corresponding pano ID for each matched GSV point was recorded.[2]

The archive first goes through an occlusion check, which verifies whether the building of interest is actually visible from the camera's position. A basic algorithm draws a straight line from the GSV point (camera location) to the centroid of the target building. If this line intersects any other building polygon in the cadastral dataset, the image is considered occluded and is excluded from the archive. This process ensures that visibility-based errors—such as interpreting a structure as absent or demolished when it is merely obscured—are minimized (Figure 4.2). These conditions and intersections are illustrated in Figures 4.3 and 4.4 in Karaköy.

[2]A Pano ID (Panorama ID) is a unique identifier used in the Google Street View system.

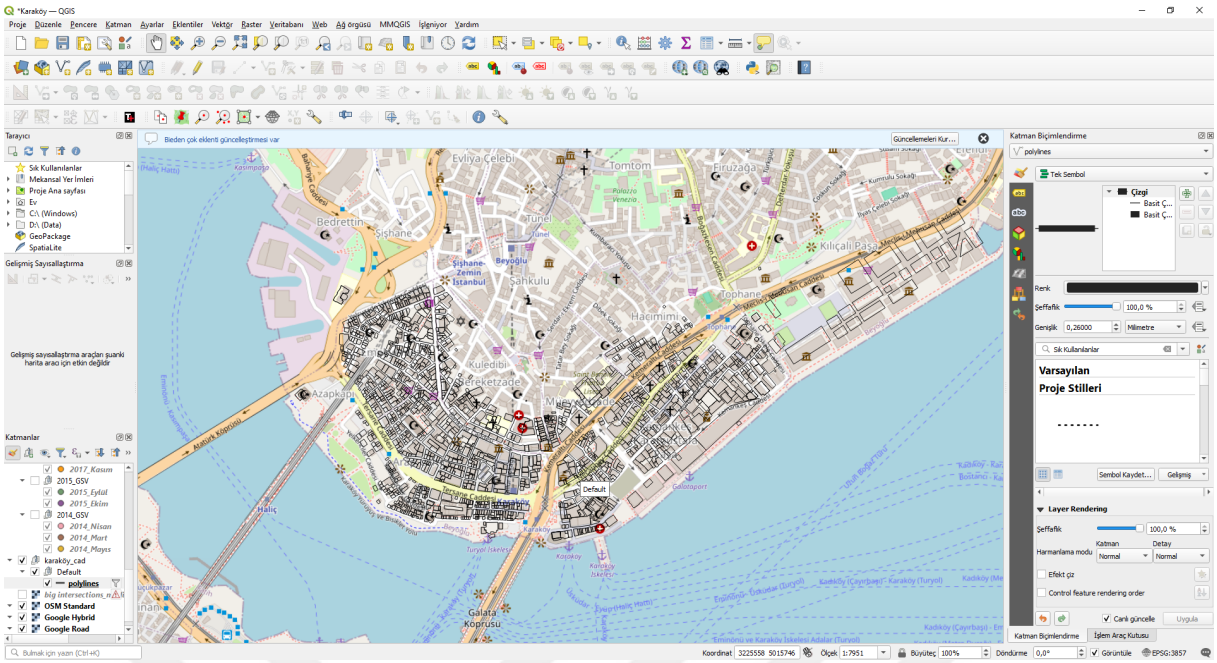


Figure 4.1: Juxtaposing the Google Maps, OSM Data, and CAD file from Municipality in QGIS.

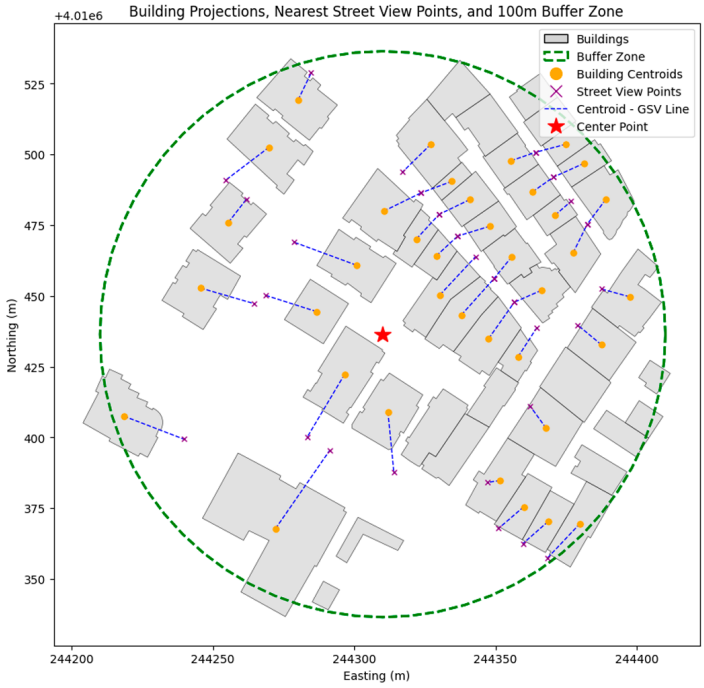


Figure 4.2: Occlusion check: intersection of line-of-sight from GSV point to building centroid

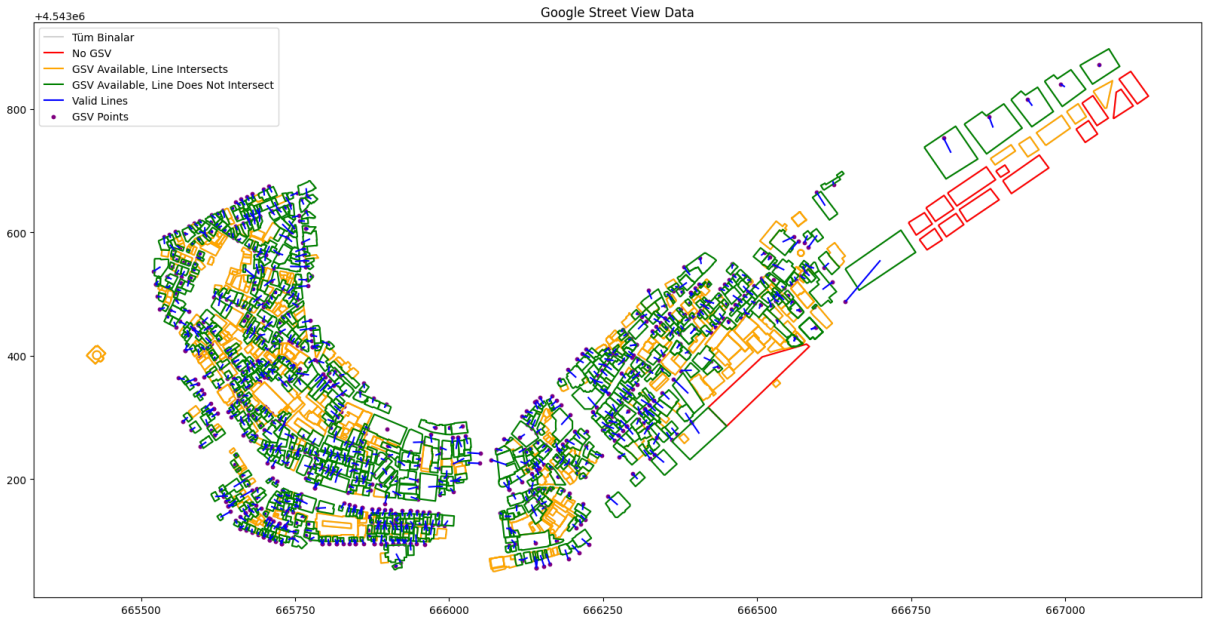


Figure 4.3: GSV points and occlusion-based exclusions based on building intersections

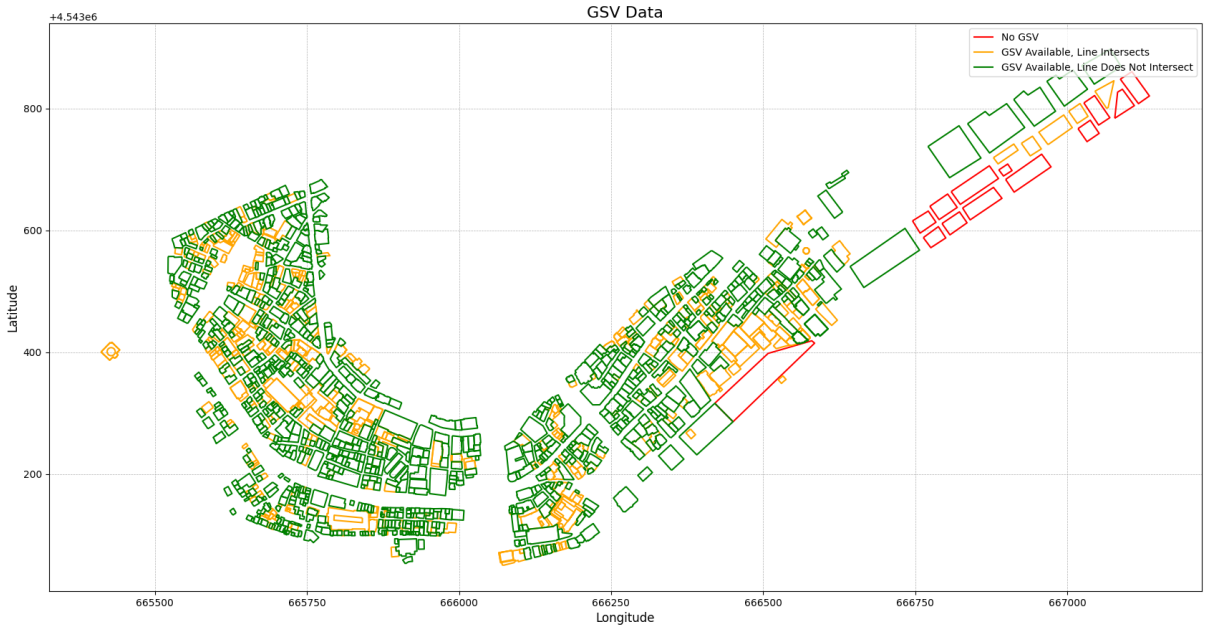


Figure 4.4: Categorization of data quality based on visibility conditions

Second challenge encountered was related to the building base layer, which was derived from a single temporal dataset. Although this dataset represents the most current state of the

area, its nature as a polygon-based representation introduced several limitations. Specifically, variations in scale, geographic positioning, and measurement accuracy across different time periods have led to inconsistencies when attempting to retrieve or align building data. However, since GSV (Google Street View) provides the most up-to-date points and imagery, the choice of using the most current dataset ensured a better match between street-level observations and the spatial data. Once the 2024 polygons were established as the origin geometry in Karaköy, subsequent GSV queries were initiated from these locations, working backward through time to retrieve historical panoramas that aligned with the most current known building extents. This anchoring approach helped maintain geometric consistency across temporal slices despite underlying morphological change.

Another challenge was the significant variation in building scale, especially in the Galataport area, where structures are substantially larger than the average building stock in Karaköy. To ensure spatial consistency and adequate facade coverage, road-based sampling was employed alongside building data. A median interval of 25 meters was selected to balance between image redundancy and data gaps (Figure 4.5). The Galataport boundary was defined as a polygon, and road data from OpenStreetMap (OSM) within this area was extracted and prepared for precise distance calculations by adjusting it to metric units. Each road polyline was exploded into segments, and bearings were calculated to align GSV capture directions with the road axis. Perpendicular headings (90° and 270°) were assigned to approximate the orientation of adjacent facades. A fixed pitch of 25 was used to ensure consistent facade visibility across samples.

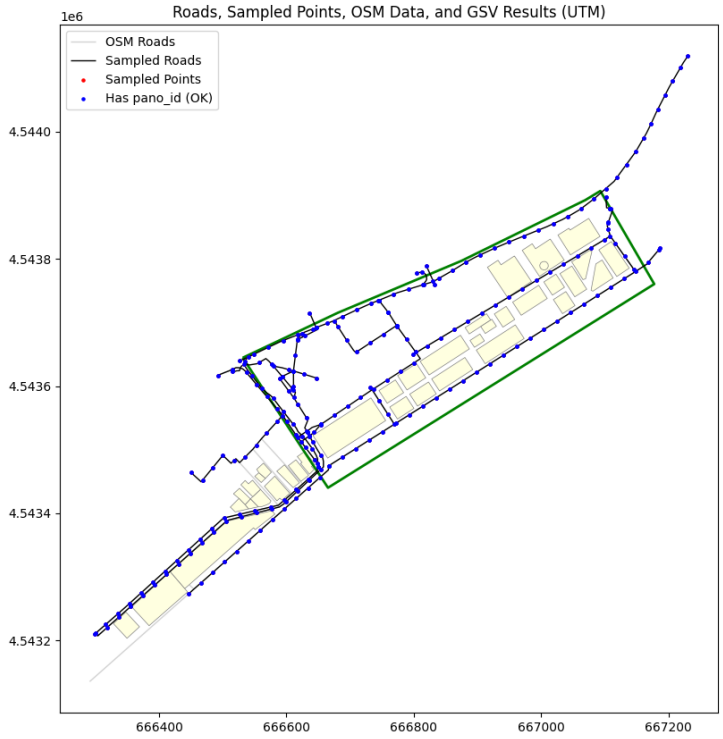


Figure 4.5: GSV points every 15 meters around Galataport

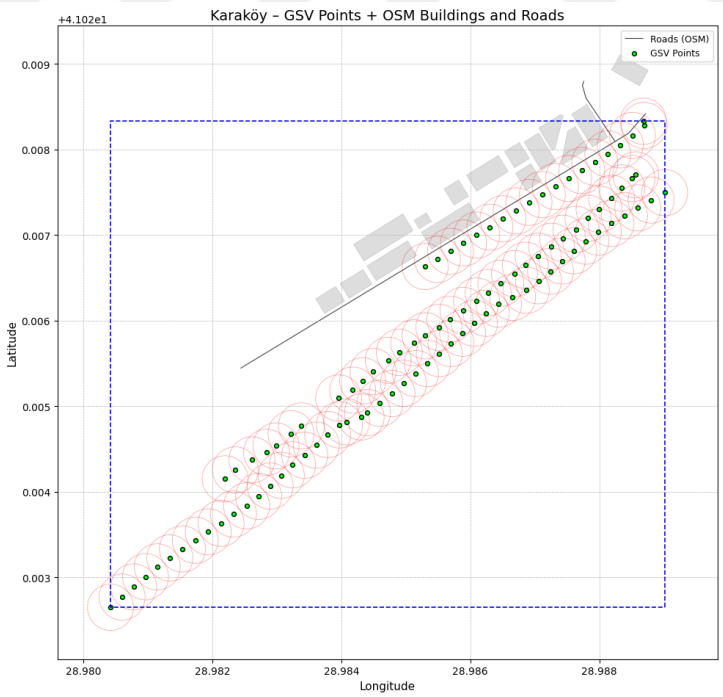


Figure 4.6: GSV points every 15 meters on shoreline

Lastly, to extract coastline data, an inverse workflow was applied. Specifically, two polylines representing Google Street View (GSV) traces were manually digitized in QGIS based on visual references from existing Google Maps imagery. To approximate the spatial alignment between the datasets, 25-meter-interval dummy points were generated, and the nearest points on each polyline were matched accordingly in Figure 4.6. In order to simulate the GSV viewing direction, a fixed bearing of 330 degrees was assigned to each point along the polylines in Figure 4.7. Subsequently, an additional set of dummy points was created to identify locations where the two polylines spatially intersect or closely overlap. At these locations, a buffer with a 5-meter radius was applied to ensure that both lines refer to the same physical area. This methodology was essential, as the polylines were derived from separate coastline surveys conducted at different times and represented slightly offset but parallel trajectories. The designed workflow thus ensures spatial consistency and alignment across temporally distinct yet geographically corresponding shoreline data.

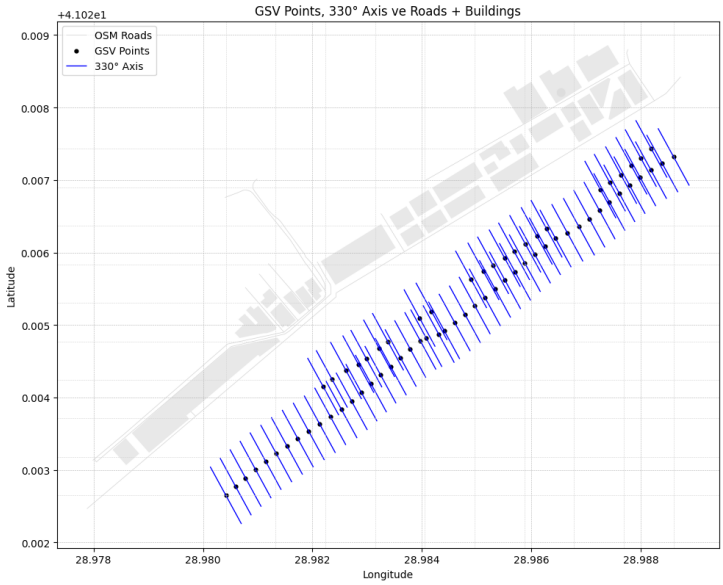


Figure 4.7: Axis assigned to each point to match GSV images

After this step, the remaining clear-view images were grouped into temporal grids, with each grid containing multi-year snapshots of the same building facade taken from a consistent angle (Figure 4.8). For each facade, the most recent image—typically from 2024—was used as the reference, and its pano ID served as the temporal anchor. All other images in the grid were

aligned to match its heading, ensuring consistent viewpoints across years. These grids were then dynamically loaded by a custom script and fed as tensors into a multimodal foundation model.



Figure 4.8: Individual street views grouped into a temporal grid as ping.

Multimodal foundation models provide a flexible base for downstream applications by learning generalized representations across different modalities. In this study, a vision-language pipeline was implemented using OpenAI's API services, specifically leveraging the capabilities of ChatGPT-4.1. This architecture integrates a visual encoder (VLM) with a large language model (LLM), enabling the model to process visual input alongside textual reasoning. The VLM component is responsible for structured description, semantic interpretation, and cross-temporal comparison, making it suitable for detecting and classifying facade-level urban changes. The LLM complements this by interpreting the VLM's outputs in context, generating coherent and structured explanations. Together, this multimodal framework supports complex

reasoning across image sequences collected from different time points. In this implementation, OpenAI's multimodal interface was used to generate structured JSON-based outputs, which allow for both detailed scene description and temporally grounded classification of transformation types. [3]

Prompt engineering regulates the model's output into two phases that approximate *denotation–connotation distinction* while maintaining machine readability. The first phase instructs the model to describe physical entities visible in each sub-image; the second phase asks what these entities potentially signify about structure, function, or socio-cultural identity. Both answers are constrained by a JSON schema bearing two keys. The key image descriptions hold an ordered list of sentences, each tied to a time-indexed sub-image, and the key description contains a collage-level synthesis. Because the model is compelled to speak in JSON, the language is converted into data that can be queried, filtered, and joined without fragile natural-language parsing. Structured language here is not a convenience but a methodological safeguard: the schema forces semantic alignment between human enquiry and machine output, reducing the hermeneutic gap that normally accompanies large-language-model (LLM) text. The prompts employ general evaluative language rather than domain-specific terminology, reflecting an intentional effort to preserve model-agnostic interpretability; the influence of prompt phrasing on generated outputs is acknowledged.[4]

To operationalize this typology and enable interpretable detection of transformation types, the framework employs Chain-of-Thought reasoning in combination with structured JSON outputs derived from diachronic visual captions. The structured responses obtained through the

[3] Vision-Language Models (VLMs) have emerged at the intersection of advances in computer vision and natural language processing. Initially, Convolutional Neural Networks (CNNs) enabled hierarchical visual feature extraction, while transformer-based architectures introduced self-attention mechanisms that improved contextual modeling in language. The adaptation of these mechanisms to visual data through Vision Transformers (ViTs) allowed for global spatial reasoning. In parallel, large-scale language models such as BERT and GPT-3 advanced semantic understanding and generalization. VLMs synthesize these developments by aligning visual and textual representations in a shared embedding space, enabling multimodal reasoning tasks such as captioning, retrieval, and temporal scene interpretation.

[3] Prompt engineering is the process of designing input prompts to guide and optimize the outputs of large language models (LLMs). Rather than treating prompting as a trial-and-error interaction, recent studies frame it as a systematic methodology for task specification and model alignment (Boonstra, 2025; Giray, 2023; White et al., 2023). It involves formulating inputs that clarify intent, structure reasoning, and reduce ambiguity, effectively functioning as an interface layer between user goals and model behavior.

[4] A sample JSON-formatted prompt is included in the Appendix.

JSON schema enable a series of follow-up queries that go beyond superficial inspection. These queries operate at the level of semantic reasoning, interrogating the structured data to extract specific indicators of transformation. For instance, one query may employ regular expressions to identify all image descriptions referencing materials such as ”composite cladding,” accounting for lexical variation. Another may detect shifts in signage, such as the appearance or removal of bilingual texts, between consecutive timestamps. Each targeted query produces a refined subset of the original data, which is then reintegrated into the main dataset. Through this iterative enrichment process, the dataset evolves from a static collection of pixel-based observations into a multi-layered analytical archive.

Previous studies have typically approached urban transformation through perceptual scoring models or binary change detection, often treating change as a continuous or discrete signal without explicitly differentiating its material or functional dimensions. In contrast, the proposed classification of urban transformation into *structure, facade, and function* reflects distinct levels of spatial intervention, each associated with different modes of altering the built environment. Structural changes often involve substantial material modifications that redefine the form, massing, or spatial configuration of buildings. Facade-level interventions are typically limited to surface treatments such as cladding, color, or signage adjustments, which influence the visual perception of space without modifying its core physical attributes. Functional transformations refer to shifts in use or program, such as a change from residential to commercial, which may alter the socio-economic role of a building without changing its physical structure. While existing approaches frequently represent these differences through perceptual scoring systems, such models may overlook the varying intensities and implications of each transformation type. In urban contexts shaped by capital-intensive development and globalization, it is essential to define and calibrate the semantic reasoning process in relation to the scale and depth of intervention. Doing so enables a more analytically robust interpretation of change and supports the development of computational models that can distinguish between superficial renewal and substantive reconfiguration.

Table 4.1: Structure and Interpretation Rules of JSON Prompt

Field	**Description**
image_descriptions	A chronological list of sub-image descriptions, each formatted as "`<YEAR>: <description>`". Each entry corresponds to a timestamped sub-image in the collage. Only images with visible timestamps should be described.
description	A detailed analysis of how the central building changes over time, synthesizing information from the image descriptions. Interpretation must follow chronological order.
transformation_type	A list of transformation events, where each item includes a *type* and a corresponding *year*.
type values	One of: *Facade*, *Structure*, *Function*, *No Building*, or *Insufficient Data*.
Evaluation criteria	Interpret only if more than 50% of the central building is visible. Camera angles and lighting may vary. Disregard fences, bridges, vegetation, and adjacent or side buildings.
What not to interpret as transformations	Minor changes such as shutter position (open/closed), graffiti, color fading, or natural weathering should not be considered as valid indicators of transformation.
Transformation definitions	• *No Building*: No visible structure in any image. • *Insufficient Data*: Only one timestamped image is available. • *Facade*: Alterations in surface material, color, signage, or window structure (excluding natural aging). • *Structure*: Physical changes to form such as demolition, rebuilding, or additional floors. • *Function*: Apparent change in building use, inferred from signage or layout; shutter closures or graffiti do not qualify.

4.2 Layering with Signification

This study adopts a multilayered analytical framework based on Denis Wood's epistemology of mapping. Within this perspective, spatial representations do not simply describe the urban environment but actively constitute processes of meaning-making. By incorporating

multimodal data sources, including AI-generated visual indicators (pings), temporal clustering methods, and the context, this research establishes a structured interpretive approach organized into four analytical layers: *Elemental, Systemic, Synthetic, and Presentational*. Each of these layers corresponds to a distinct stage, progressing from the identification of micro-scale material changes to the analysis of broader patterns of urban transformation, culminating in their visual and discursive representation. Within this framework, *pings* refer to algorithmically identified visual anomalies—such as signage alterations, facade changes, or structural disappearances—*that function as the smallest observable units of spatial transformation.* In simplified terms, the Elemental layer focuses on what is observed, the Systemic layer on how these observations are spatially and temporally organized, the Synthetic layer on what these patterns signify in a broader socio-political context, and the Presentational layer on how such interpretations are ultimately communicated and visualized.

This layered framework is not only a descriptive model of how spatial knowledge unfolds, but also a prescriptive framework for how mapping practices themselves can be designed to reflect and intervene in that process. Significantly, if spatial meaning emerges through processes of layered perception and interpretation, then mapping should not be limited merely to representing spatial outcomes. Rather, by consciously incorporating these layered processes into the very methodology of cartographic production, *mapping itself can become a critical practice.* In this approach, cartography moves beyond the passive portrayal of spatial realities and engages directly with the politics inherent in representation. When pings begin to cluster spatially or temporally—especially within morphologically or functionally coherent zones—they may coalesce into signals, representing systemic patterns that suggest higher-order processes at work. Thus, mapping is reframed as a reflexive and interpretive process, wherein every visual and semantic decision reveals the constructed and situated character of spatial knowledge. The objective of this approach is to identify and visualize urban transformations while simultaneously engaging in a critical examination of the methodological decisions and socio-technical conditions that shape what becomes visible and what remains obscured. Consequently, the multilayered methodology presented here positions mapping as an epistemologically informed and critically performative mode of inquiry, one that brings to light the intertwined relationships

among computational vision, spatial politics, and the active construction of urban knowledge.

4.2.1 Elemental Layer

In the analytical architecture proposed here, algorithmically structured image outputs are conceptualized as *pings—elemental indicators* of micro-scale spatial change. These indicators are not interpretive symbols in themselves, but computational artefacts (outputs) derived from AI-based models that extract visual discontinuities across time-indexed imagery or metadata-driven traces, such as shifts in storefront names, signage alterations, or facade reconfigurations. The designation of these outputs as pings does not imply inherent significance; rather, their epistemological activation is contingent upon a human-constructed interpretive schema. They function as raw visual cues; dispersed fragments within a broader semiotic and spatial field, whose meaning emerges only when situated in relation to other indicators and temporal frames.

These pings constitute the empirical foundation of the Elemental layer in Wood's stratified model of mapping in 4.9. At this initial stage of analysis, the primary objective is strictly descriptive: to identify, document, and systematically archive discrete changes that occur within the district. To operationalize this detection, transformation types were annotated using a predefined ontology set, formalized across both textual and visual components of the analysis. This ontology was structured around a typology of five distinct transformation categories: Facade, Structure, Function, No Building, and Insufficient Data (Figure 4.9). Each transformation event was classified according to visible changes detected in the central building across timestamped sub-images.

The classification framework delineated facade transformations as surface-level modifications—such as changes in signage, cladding, or fenestration—that do not alter the building's volumetric configuration. Structure referred to substantial modifications in physical form, including demolition, vertical extensions, or reconstructions. Function captured inferred shifts in building use, typically indicated by the appearance or replacement of commercial signage, internal spatial arrangements, or facade articulations that signal programmatic change. Instances where only a single timestamped image was available were labeled Insufficient Data, while

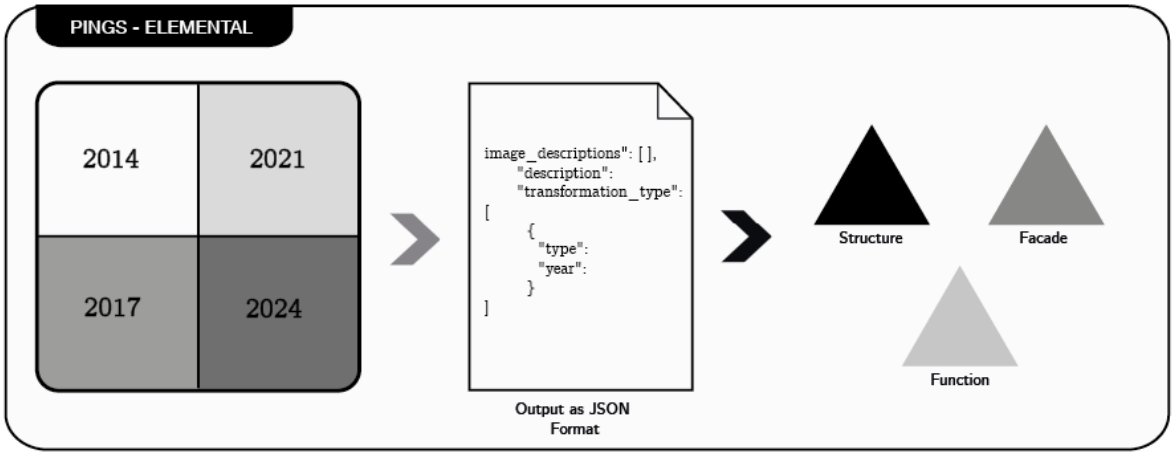


Figure 4.9: Framework of Elemental Layer

scenes lacking a discernible central structure altogether were designated as No Building.

In the visual representation of the Elemental layer, each transformation event was denoted using a triangular symbol in Figure 4.10. This choice was intentional: while point markers are often employed in later layers—such as the Systemic layer—to represent aggregated transformation clusters, the triangle was adopted here to emphasize the individuality and directional specificity of each ping. In cases where multiple transformation types co-occurred within the same spatial frame, a representational hierarchy was established to resolve overlaps: Structure was prioritized over Facade, which in turn took precedence over Function. This ordering reflects the assumed material and spatial impact of each category, with structural changes considered the most enduring and morphologically significant. Although one could argue that functional or facade-related alterations may carry greater spatial or social implications in specific contexts, the hierarchy applied here reflects a methodological emphasis on visibly legible, physically grounded modifications within the built environment. While this visual encoding facilitates the granular identification of transformation types, the analytical capacity of the Elemental layer remains limited in isolation.

However, the analytical significance of a ping does not reside solely in its appearance within the Elemental layer. The epistemic value of such micro-indicators depends on their subsequent organization, contextualization, and interpretation through the higher-order layers

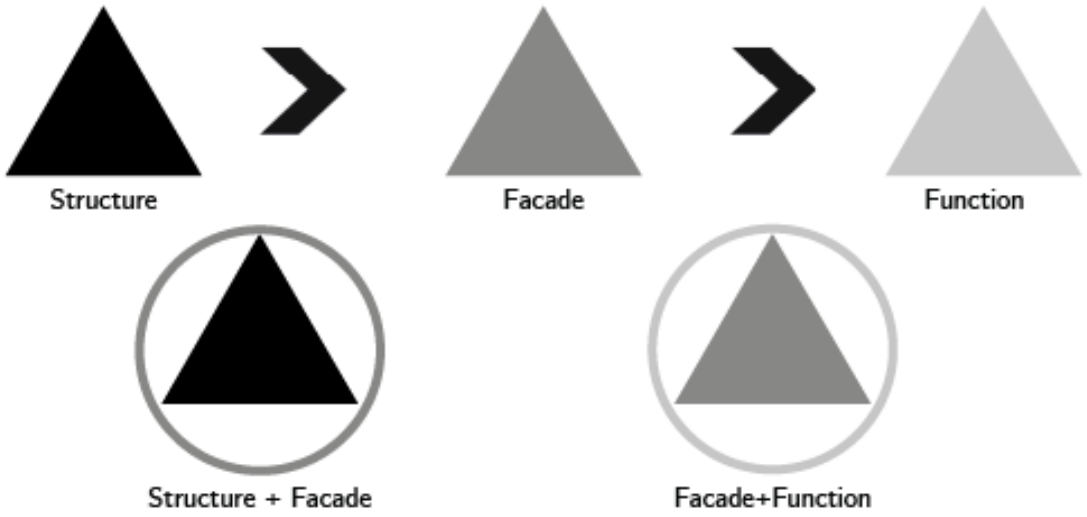


Figure 4.10: Visual encoding of transformation types within the Elemental layer using triangular symbols.

of the framework. Without the structuring logic of the Systemic layer, which assembles pings into spatial and temporal patterns, or the interpretive abstraction of the Synthetic layer, which links those patterns to broader socio-political processes, pings risk remaining analytically inert.

4.2.2 Systemic Layer

The Systemic layer in this study is constructed through a dual-pathway analytical approach that integrates both visual interpretation and unsupervised computational clustering, each addressing different dimensions of spatial reasoning (Figures 4.11 and 4.11). The first approach relies on the diachronic mapping of pings—elemental indicators of spatial transformation—across time-indexed layers. By distributing these events visually on a temporally sensitive map interface, it becomes possible to perceive local concentrations, sequential alignments, or absences through interpretive observation. In this context, the map does not merely function as a representational backdrop but becomes a semi-analytic surface that invites spatial diagnosis through proximity, co-occurrence, and morphological context. Although this technique relies on visual perception and thus verges on the Presentational layer, it is here intentionally positioned

within the Systemic level to emphasize its role in pattern detection, rather than communication alone. This visual pathway is particularly effective in cases of limited or thematically bounded data, where human cognition can still reliably interpret spatial densities and temporal rhythms without algorithmic assistance.

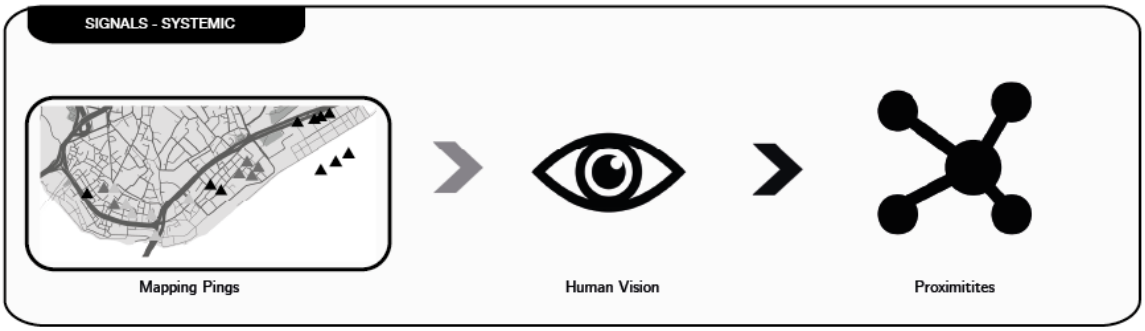


Figure 4.11: Visual interpretation of transformation events through diachronic mapping of pings, enabling human perception of spatial proximities and emergent systemic patterns.

However, as the dataset expands—both in volume and complexity—visual perception reaches its cognitive limits. As noted in the second panel (Figure 4.12), the interpretive capacity of human vision is insufficient to resolve thousands of overlapping pings distributed across heterogeneous urban morphologies. In response, a second methodological path is employed: unsupervised clustering through Hierarchical Density-Based Spatial Clustering of Applications

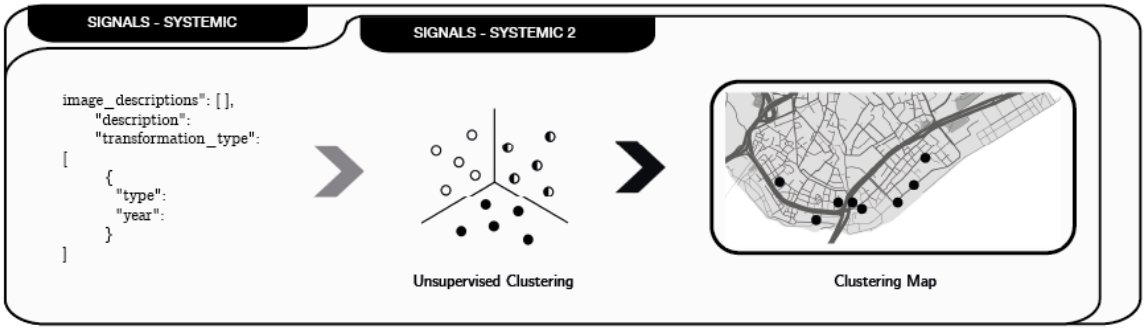


Figure 4.12: Unsupervised clustering of transformation data using HDBSCAN, producing data-driven groupings of events visualized as transformation clusters on the urban map.

(HDBSCAN) with Noise.[5] This algorithm identifies event clusters based on local density variations and generates probabilistic outputs that differentiate between core clusters, noise, and border regions. Crucially, this process does not depend on pre-specified cluster counts or rigid parameterization, allowing transformation patterns to emerge organically from the data itself. This is especially important in the context of urban environments like Karaköy, where spatial logic does not conform to regular grid patterns or evenly distributed functions. Binary flagging is applied to structure, facade, and function attributes to allow multi-label representation within the clustering input, as HDBSCAN operates exclusively on numerical feature spaces.

While the Systemic layer establishes the spatial and temporal coherence of transformation events, it does not, in itself, account for their broader significance. The identification of clusters, whether through visual reasoning or unsupervised learning, reveals where and when transformation occurs and how events are spatially configured, but not necessarily why such patterns emerge or what they imply within larger socio-political or economic trajectories. To engage these questions, the analysis must ascend to the Synthetic layer, where systemic formations are recontextualized within broader interpretive frameworks. This shift entails a movement from spatial pattern recognition to conceptual abstraction, wherein observed regularities are interpreted in relation to urban development logics, planning interventions, policy histories, or shifting modes of socio-economic organization. It is at this level that transformation clusters acquire explanatory value, serving not only as spatial phenomena but as indicators of structural processes such as displacement, gentrification, privatization, or infrastructural realignment that shape the production of urban space over time.

4.2.3 Synthetic Layer

The Synthetic layer constitutes a higher-order interpretive stage in which spatial patterns identified in the Systemic layer are abstracted and re-embedded within broader socio-political,

[5]HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is an unsupervised machine learning algorithm used to identify clusters in data without requiring labeled inputs or a predetermined number of clusters. It extends traditional density-based clustering by building a hierarchy of clusters based on local point density and extracting the most stable clusters through a condensed tree structure. This allows it to detect clusters of varying shapes and densities, while also distinguishing noise points that do not belong to any cluster.

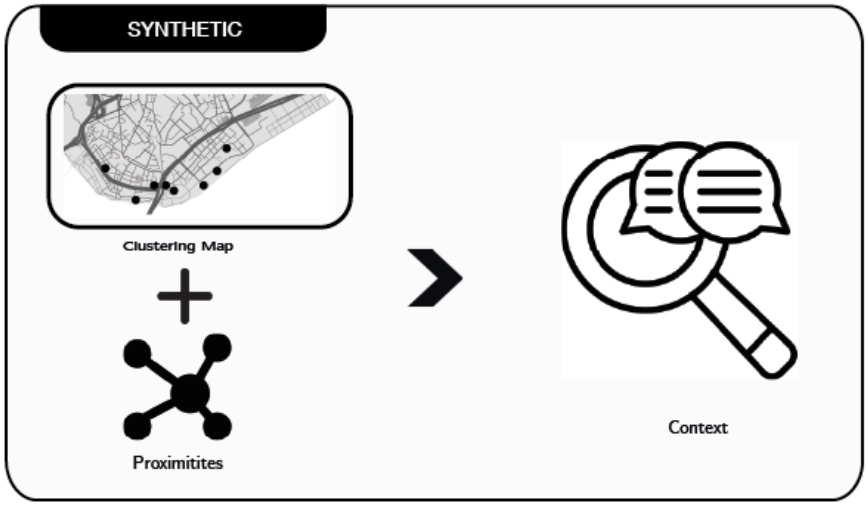


Figure 4.13: Schematic of the Synthetic layer, where spatial clusters are linked with contextual data to support interpretive analysis of urban transformation.

economic, and historical contexts in Figure 4.13. At this level, transformation clusters are no longer treated merely as spatial phenomena or geometric densities, but as manifestations of larger structural processes. Recurrent modifications, such as consecutive storefront closures, homogeneous facade treatments, or vertical building extensions, are interpreted as indicative of systemic shifts including gentrification, speculative redevelopment, commercial displacement, or changes in regulatory pressure. In this context, spatial formations acquire explanatory value, linking the material footprint of transformation to deeper urban logics that operate across institutional, economic, and temporal scales.

To facilitate this process, spatial clusters are intersected with a range of auxiliary data layers, including zoning maps, cadastral boundaries, land-use designations, tourism metrics, historical preservation records, and archival planning documents. This multimodal integration enables the identification of correlations between localized spatial change and broader governance regimes or market dynamics. Importantly, the Synthetic layer also accommodates outlier events, pings that diverge from the dominant pattern within a cluster, as potential indicators of unexpected changes, local opposition, or areas with limited policy enforcement. These interpretive readings set the stage for the Presentational layer, wherein the findings are translated into communicative visual formats. In this final layer, emphasis shifts from analysis to representation: from the conceptual abstraction of transformation to the epistemological framing of

how such change is made visible, narratable, and actionable within public, institutional, and academic contexts.

4.2.4 Presentational Layer

The Presentational Layer constitutes the final stage of the visual-analytical workflow, serving as a synthetic platform in which outputs from the Elemental, Systemic, and Synthetic Layers are juxtaposed to reveal the multi-scalar and multi-temporal structure of urban transformation. This layer does not merely compile the outputs but integrates them through spatial superposition, enabling the concurrent reading of raw signals, clustered patterns, and interpretive overlays. All relevant visual materials, including GSV coverages, transformation signals, clustering outputs, and annotated interpretations, are rendered as translucent sheets to allow for physical alignment and comparative inspection in Figure 4.14. The entire assembly is designed for tactile manipulation, foregrounding the spatial dynamics of transformation through layered transparency. The base map is rendered using Nolli Map in a muted alt-gray tone to preserve scale readability and typological clarity in a 2D format. The use of Nolli Map as a foundational cartographic substrate ensures that fine-grained morphological distinctions remain legible even under multiple overlaid surfaces, making the Presentational Layer both a composite analytical device and a material representation of the study's layered methodology.

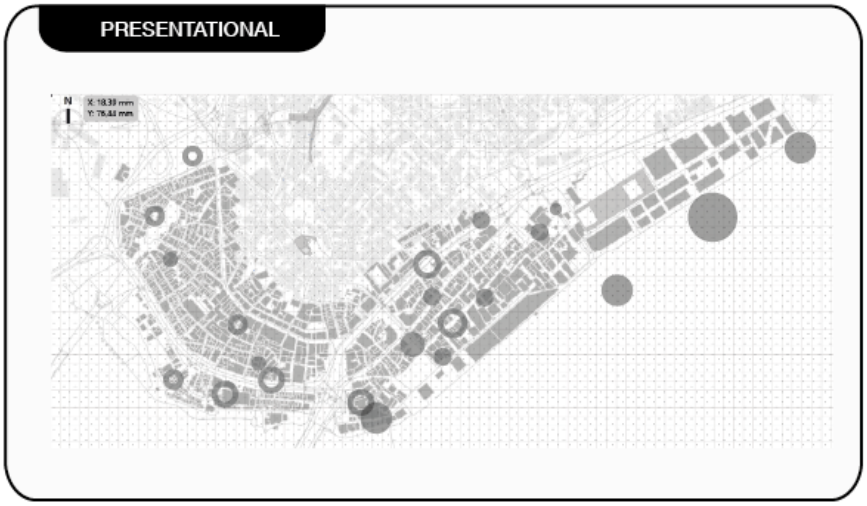


Figure 4.14: Presentational Layer showing the cumulative GSV coverage of all years over a NolliMap basemap, with overlaid Boolnaz Signals and systemic inferences from clustering analyses.

4.3 Results and Discussion

4.3.1 Implementation Outcomes from the Karaköy

The visual transformation detections in the Karaköy district are analyzed in this section. Collages that indicate transformation, pings are classified for each year individually and grouped under three primary categories: structure, facade, function, no building, and insufficient data. These signals are interpreted through their spatial clustering and temporal progression within a four-layered analytical framework: Elemental, Systemic, Synthetic, and Presentational.

The outputs generated by the model are not confined to quantitative comparisons alone; rather, they facilitate a multi-dimensional interpretation of each year's distinct data configuration. This includes variation in representational density, typological distribution of transformation categories, and their spatial articulation within the urban fabric. Notably, the year 2024

constitutes a methodologically significant benchmark, distinguished by its high-resolution coverage and the simultaneous visibility of all transformation types.

The comparative assessment of image availability and transformation-type distribution, as illustrated in Figures 4.15, 4.16 and 4.17, indicates a strong positive correlation between the volume of Street View data and the model's efficacy in detecting and categorizing urban transformations. Years characterized by high image density, most notably 2024, followed by 2018 and 2017, correspond with an increase in transformation outputs, particularly within the facade and function categories. Notably, in a small subset of the 2018 and 2024 datasets, multiple images were captured for the same location within the same year. While such cases are relatively infrequent, they occasionally produce intra-annual variation in detected transformations, which may introduce minor redundancies or inflated change signals in localized areas.

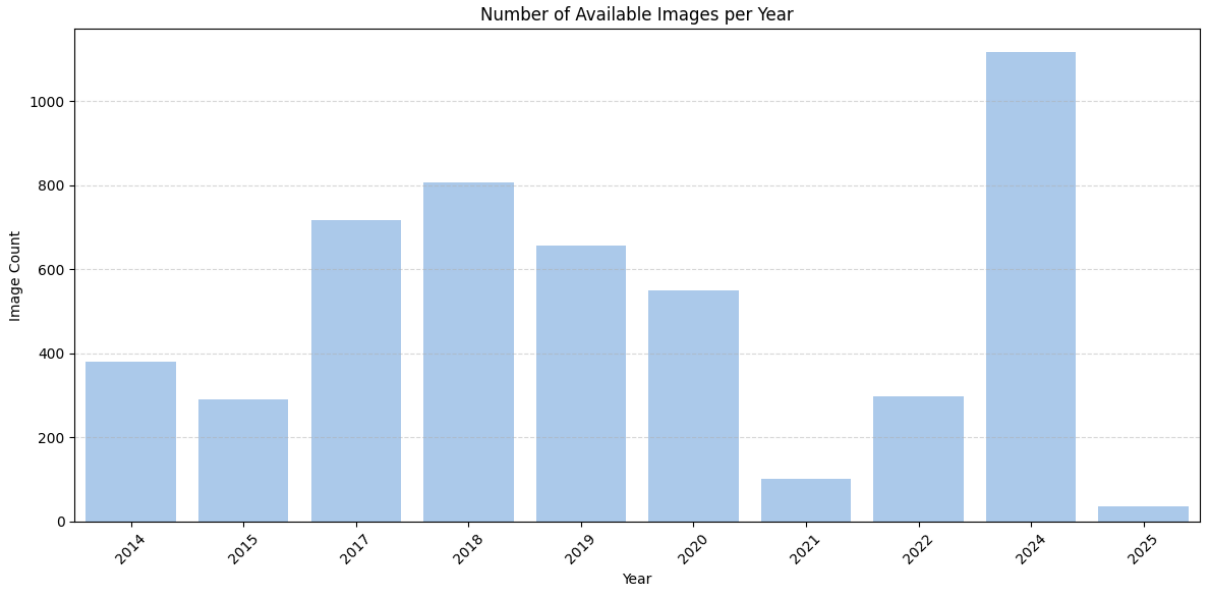


Figure 4.15: Annual distribution of Google Street View images used in the study, showing variation in availability over time.



Figure 4.16: Distribution of Detected Transformations and Available Images per Year

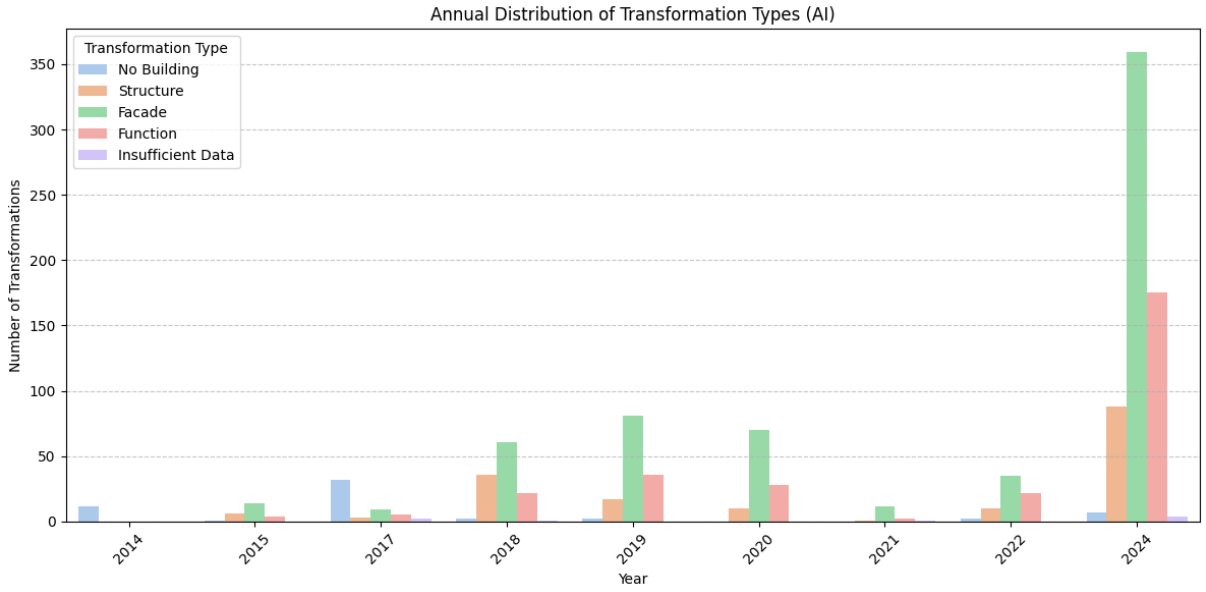


Figure 4.17: Annual distribution of transformation types detected by the AI model between 2014 and 2024.

The year 2024, supported by approximately 900 georeferenced images, exhibits a substantial escalation across all transformation types, thereby affirming its analytical role as the origin year for temporal benchmarking. In contrast, 2021 reflects a pronounced data void, with limited

imagery resulting in significantly reduced detection rates and exposing the model's sensitivity to representational discontinuities. The persistent predominance of facade-related transformations across all years suggests a systematic model bias toward surface-level visual cues such as signage, materiality, and color alterations. Structure and function transformations, on the other hand, demonstrate greater temporal variability, aligning with periods of intensive construction and redevelopment. A longitudinal analysis of urban transformation was conducted by comparing pings and their spatial clusters across all ten available years (2014–2015, 2017–2020, and 2022–2024) through a series of GSV coverages, enabling a spatially integrated and temporally diachronic interpretation of change signals and pings.

As one of the earliest temporal layers of the study, the 2014 dataset presents limited spatial coverage and continuity in Google Street View imagery. Nonetheless, it provides analytically valuable initial signals for interpreting subsequent urban transformations. The Figure 4.18 reveals three major zones of spatial concentration. First, the Galataport waterfront, particularly along the Tophane shoreline, contains pre-construction imagery that enables before-and-after comparisons for later large-scale interventions. These early visual records serve as a temporal baseline for detecting structural and functional transformations in subsequent years. Second, the image sequences along Meclis-i Mebusan and Tophane-İskele Streets indicate a prioritization of primary transportation corridors during data collection. In locations where road-aligned and building-centered image acquisition strategies intersect, artificial density increases appear in the GSV coverage, suggesting methodological overlap and resulting spatial intensification. Third, the intersection of Necatibey and Maliye Streets, surrounded by administrative and governmental buildings, including the Tax Inspection Board, unexpectedly exhibits high image density. In addition to its functional importance, the relative structural stability of this zone may render it useful as a control area for evaluating change detection in contrast to more dynamic zones. A fourth notable concentration occurs at the entrance to Tersane Street, particularly around Tünel, which acts as a threshold zone with concentrated data capture. This may result from repeated imaging at difficult-to-stabilize locations or from pings.

The year 2015 constitutes a turning point in terms of increased coverage and continuity of Street View data in Figure 4.19. On the eastern side of the study area, the dataset shifts from

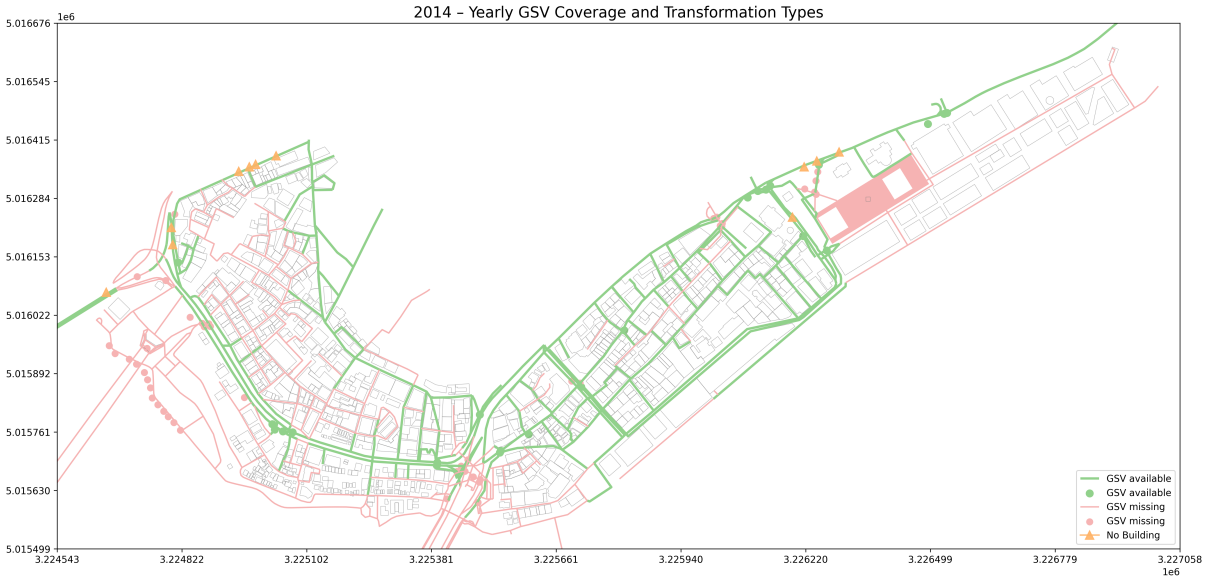


Figure 4.18: Yearly GSV Coverage and Transformation Types in 2014. Green indicates available GSV data, red indicates missing data, and orange triangles represent no-building areas.

a fragmented, arterial-focused structure toward a more continuous visual texture that includes side streets. In this period, the area around Mumhane Street is represented with a higher number of images, and early transformation signals are observed, primarily categorized as facade and function changes. These include surface-level alterations and changes in use. Meclis-i Mebusan Street also shows an increase in data density compared to the previous year. Okçu Musa Street and surrounding secondary roads are included in the dataset, where both facade and function transformations are detected. On the western side, although main roads such as Tersane Street and Bankalar Street are covered, the adjacent side streets remain unrepresented. In particular, streets such as Fermeneciler Street show gaps in visual data, and instances of "No Building" and "Insufficient Data" classifications are recorded in these areas. Additionally, structural transformation pings are recorded around Kemankeş Street, linked to the construction of large hotel chains such as JW Marriott, marking the first signs of high-impact urban redevelopment in the area.

In the absence of any available Street View imagery for the year 2016, the analysis resumes in 2017 with a rebalancing of spatial coverage. During this year, focal image clusters appear for the first time on the western side of the study area, particularly within the side-street network surrounding Tersane Street in Figure 4.21. These data points enable the identification

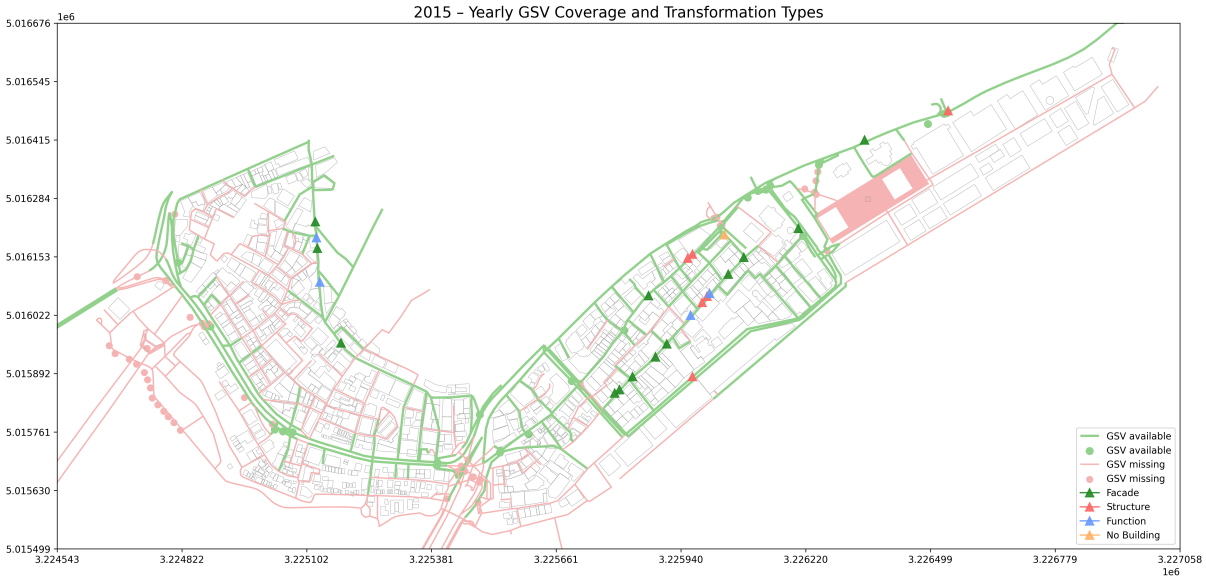


Figure 4.19: Yearly GSV Coverage and Transformation Types in 2015. In the transformation map, all markers are shown as triangles, with their colors representing different types of change: blue indicates functional transformation, green indicates façade transformation, red indicates structural transformation, orange represents building removal, and purple denotes cases with insufficient data.

of new types of transformation signals. On the eastern side, Meclis-i Mebusan Street, which had shown consistent data coverage in previous years, is not represented in 2017. In contrast, Mumhane Street and Necati Bey Street remain visible, though the number of images is reduced compared to earlier periods. Notably, the Mumhane corridor continues to exhibit a long visual sequence, uninterrupted along its central spine, allowing for the detection of sustained facade and function transformations across multiple building frontages (Figure 4.22 and 4.21. On the western side, Tersane Street is captured over a longer segment, yet the data is organized into non-continuous, point-based clusters. Streets such as Perşembe Pazarı, Söğüt, Billur, and Ergüvan exhibit concentrated pings, marking the first time that inner-street morphologies of the western district are visually recorded.

In some cases, such as along Fermeneciler and Makaracılar Streets, imagery is available, but no building footprints are detected in 2017. This issue may stem from GPS inaccuracies in open public spaces along the waterfront, such as the Galataport area. Although municipal cadastral data, OpenStreetMap, and Google Maps datasets generally align within the QGIS interface, subtle discrepancies caused by GPS signal distortions in these open areas can lead to

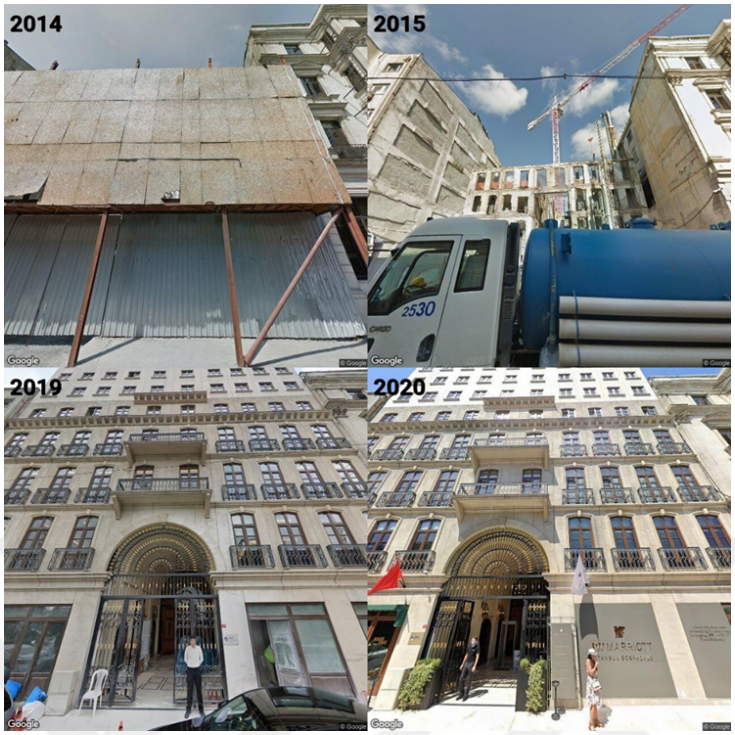


Figure 4.20: JW Marriott Istanbul Bosphorus construction phases on Kemankeş Street, one of the structure pings in 2015.

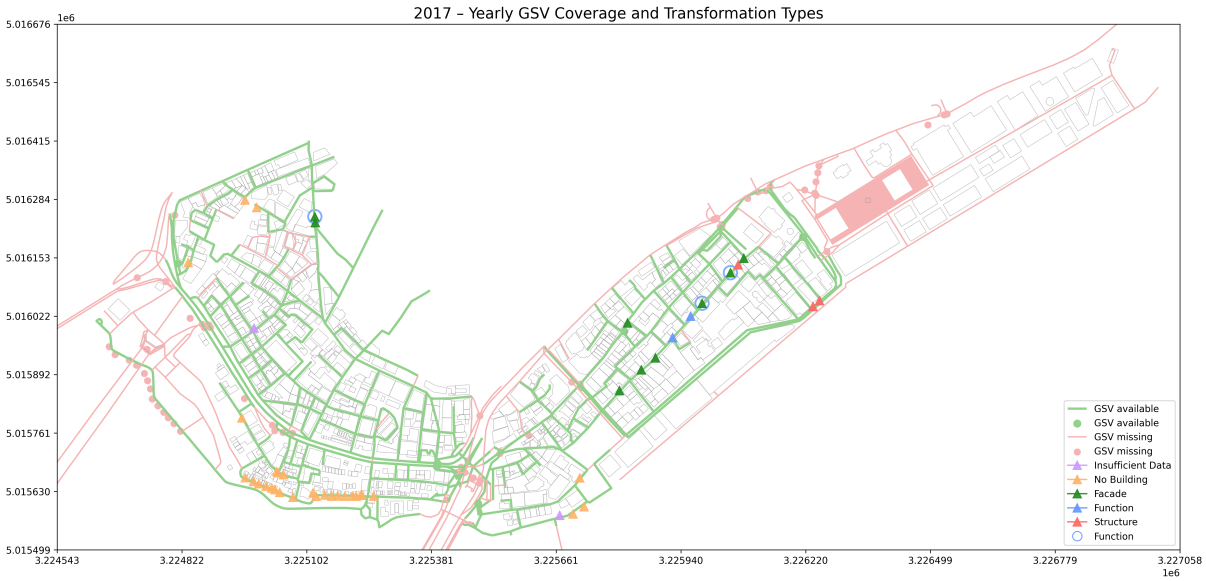


Figure 4.21: Yearly GSV Coverage and Transformation Types in 2017.



Figure 4.22: One of the depiction facade and function pings in 2017, 2018, 2019, and 2024 in Mumhane Street.

misalignments in footprint mapping. This explains why certain areas appear dense in the GSV coverage despite the absence of building detections. It should be noted that in 2017, although the viewing angle was correct and confirmed through systematic comparison with OSM data and cadastral data obtained from the municipality, a systematic error in mapping the building footprints occurred. To mitigate this issue, imagery was collected along road segments at fixed intervals of approximately 10 meters, similar to the method applied in the Galataport area. Images were captured at headings offset by 90 and 270 degrees to increase spatial coverage. Consequently, areas classified as "No Building" may appear densely represented on the GSV coverage due to this methodological sampling strategy. This factor should be considered when interpreting the data density in these zones. Finally, Mumhane Street continues to produce a high volume of transformation signals.

The year 2018 constitutes a critical juncture, marked by a sharp increase in visual data volume and a noticeable spatial clustering of specific transformation types. This period corresponds with the *increased physical visibility of the Galataport construction* and intensified interventions in the surrounding urban fabric (Figure 4.23 and 4.24). On the eastern side of the study area—including Meclis-i Mebusan, Necati Bey, Mumhane, and Kemankeş Streets—a significant concentration of data is observed. Meclis-i Mebusan Street, in particular, forms a nearly continuous capture corridor for the first time, both in terms of image quantity and transformation density. Necati Bey Street exhibits a high number of facade and function transformations, while

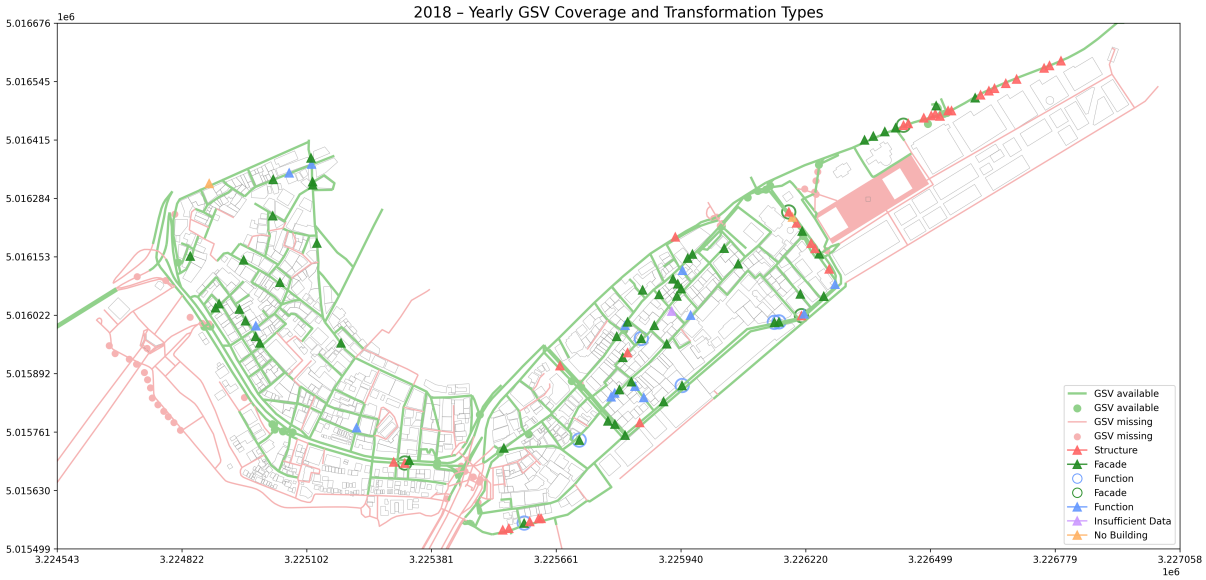


Figure 4.23: Yearly GSV Coverage and Transformation Types in 2018.

Mumhane and Kemankeş Streets show evidence of both facade and function-level changes.

By contrast, the western edge of the study area—including Tersane, Okçu Musa, and Bankalar Streets—remains more fragmented in terms of coverage and contains fewer transformation pings. Inner streets of the western section, particularly those leading inland from Tersane Avenue, are also underrepresented in 2018 despite their earlier prominence in 2017. [6]

One of the most distinct characteristics of this year remains the spatial clustering of transformation types. Structural transformations are predominantly concentrated in areas directly impacted by the Galataport construction, especially along Meclis-i Mebusan and Kılıçali Paşa Streets, where large-scale interventions into the built fabric are evident. In contrast, heritage-rich zones to the west of Galataport, such as the surroundings of Tophane Kasrı and Nusretiye Mosque, tend to exhibit more controlled or surface-level alterations, often involving enclosure, protective measures, or selective renovation rather than structural overhaul. Facade transformations in these areas are typically modest and spatially constrained. Meanwhile, functional transformations are more diffusely distributed, emerging particularly along Necati Bey Street and extending westward toward the Refik Saydam–Tersane corridor, where changes in use are

[6] In the Tersane area, specifically certain inner streets such as Abdussalah, Fütuhat, and Galata Mahkemesi, previously detected facade transformation pings have been deemed incorrect following human verification. Only structural transformations remain, and these are also likely to be misclassified. These issues will be discussed later.



Figure 4.24: Galataport construction phases, captured across nine different years.

observed even in the absence of visible architectural modification.

2018's GSV also shows that due to the methodological expansion of the dataset, which incorporates multiple image sources. As a result, the GSV coverage displays a higher visual density in Rıhtım Street, and a broader variety of transformation types is detectable. The ability to observe both facade and structure-related changes in parallel contributes to a more detailed and layered understanding of ongoing urban dynamics. In particular, many of these changes appear to coincide with the rehabilitation of previously deteriorated or abandoned urban pockets.

The year 2019 represents a transitional moment, not in the sense of construction completion, but rather in terms of how transformation signals begin to spatially diversify and extend into new zones (Figure 4.25). The Galataport project remains in its intensive construction phase during this period, and structural transformations continue to be recorded (figure 4.26, though with greater spatial variability. Instead of a singular concentration, multiple smaller clusters emerge—particularly involving facade and function changes—indicating a dispersal of interventions beyond the immediate construction core.

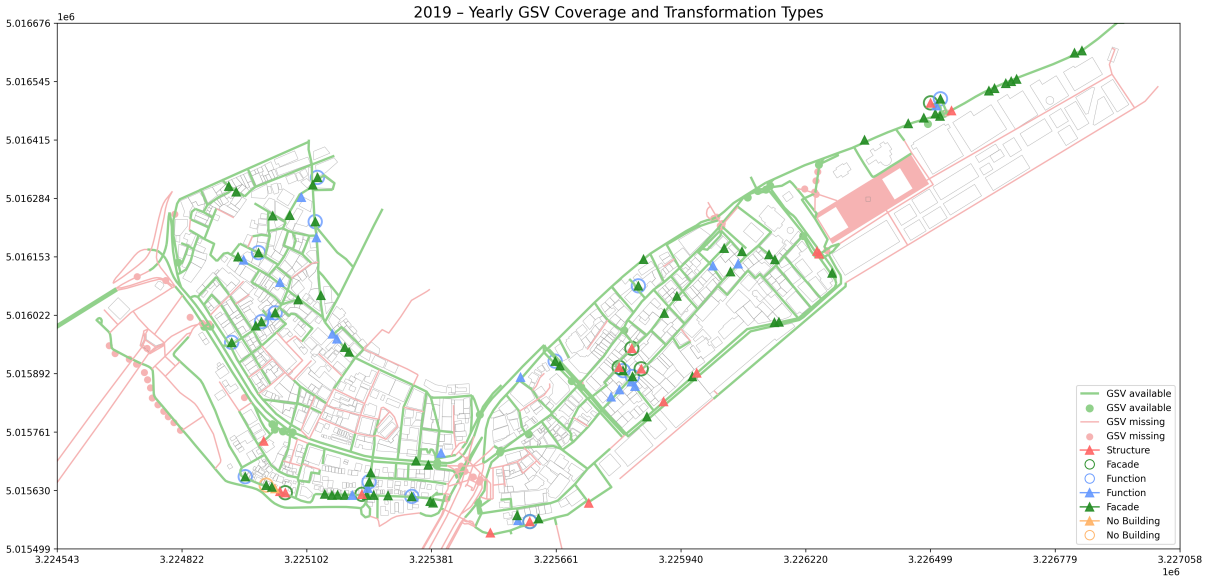


Figure 4.25: Yearly GSV Coverage and Transformation Types in 2019.

In the eastern part of Karaköy, streets adjacent to major institutional and cultural landmarks maintain data density, continuing to show facade modifications. The waterfront axis, while still under significant transformation pressure, shows signs of functional diversification in surrounding secondary corridors. On the western side, a gradual increase in data visibility occurs in the deeper segments of the urban grid, where the first signs of transformation begin to register in previously underrepresented zones.

Structural transformation remains present but no longer dominates the dataset as it did in the previous year. Instead, facade and function-related changes become more prominent, particularly in transitional corridors and secondary commercial streets. In many instances, these alterations occur simultaneously, forming hybrid transformation patterns that reflect both physical and socio-spatial shifts. Environmental obstructions such as fencing and scaffolding also continue to interfere with the clarity of visual readings, particularly in the Galataport region.

In 2020, a spatial shift in data coverage is observed, marked by a relative decline in the continuity of imagery along the eastern axis and a notable increase in focal clusters within the western segments of the study area (Figure 4.27). On the eastern side, although streets such as Tophane İskele Street and Meclis-i Mebusan Street remain partially covered, the imagery is less



Figure 4.26: Hotel construction phases are completed along Hoca Tahsin Street, with corresponding structure pings detected in 2018 and 2019.

consistent and extensive than in previous years. These areas continue to exhibit scattered transformation signals, particularly related to facade and function changes in proximity to ongoing institutional and infrastructural developments.

By contrast, the western side demonstrates an increase in image density and transformation activity, particularly in both waterfront-adjacent and more interior street segments. Streets such as Fermeneciler, Yemeniciler, and Makaracılar—all located in close proximity to the waterfront—stand out with a high frequency of facade and function transformations. These modifications typically reflect surface-level interventions and shifts in building use without structural alterations. In contrast, the Perşembe Market area, situated further inland, exhibits a different dynamic, where transformation signals emerge in a more dispersed manner, often in relation to deeper commercial activities embedded within the historical grid.

Structural transformations remain limited and spatially scattered, appearing primarily along infrastructural corridors and nodal intersections. Meanwhile, facade-related changes concentrate in secondary commercial passages, and function-related transformations are distributed across both coastal and inland areas. A recurring pattern of co-occurring facade and function

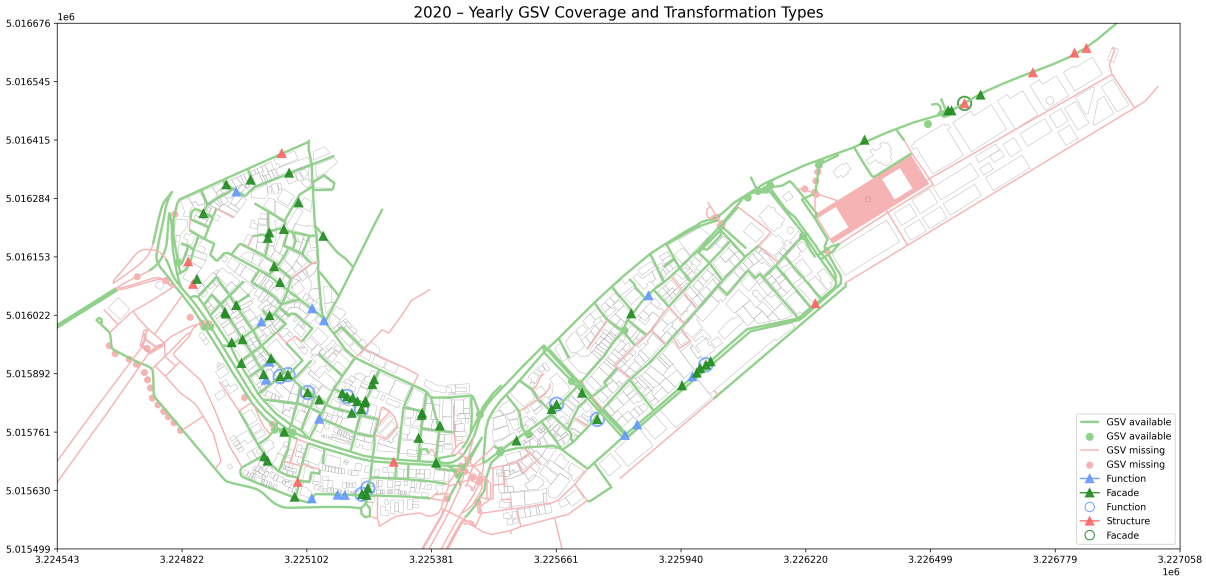


Figure 4.27: Yearly GSV Coverage and Transformation Types in 2020.

changes is observed, particularly in select zones of the western fabric, suggesting incremental yet consistent forms of socio-spatial reconfiguration.

The year 2021 is marked by a pronounced reduction in both the spatial coverage and frequency of Google Street View imagery compared to previous years (Figure 4.28). This gap may be attributed not only to logistical disruptions caused by the COVID-19 pandemic but also to the widespread closure of public spaces and a general slowdown in socioeconomic activity, which may have reduced the perceived need for documenting urban change during that period. This contraction results in a significant disruption in spatial continuity, substantially limiting the model's capacity to detect and track transformation signals across the study area. Most notably, the western portion is entirely absent from the dataset, creating a critical void in temporal representation, especially in zones that had previously exhibited concentrated transformation activity.

On the eastern side, coverage is largely limited to a single corridor along Meclis-i Mebusan Street, which serves as the primary axis of imagery. A partial and fragmented presence is also observed around the peripheries of adjacent streets, but the data is notably sparse and lacks spatial depth. Detected transformation signals in this area primarily pertain to facade changes, particularly along the eastern boundary of the Galataport construction zone. A minimal number

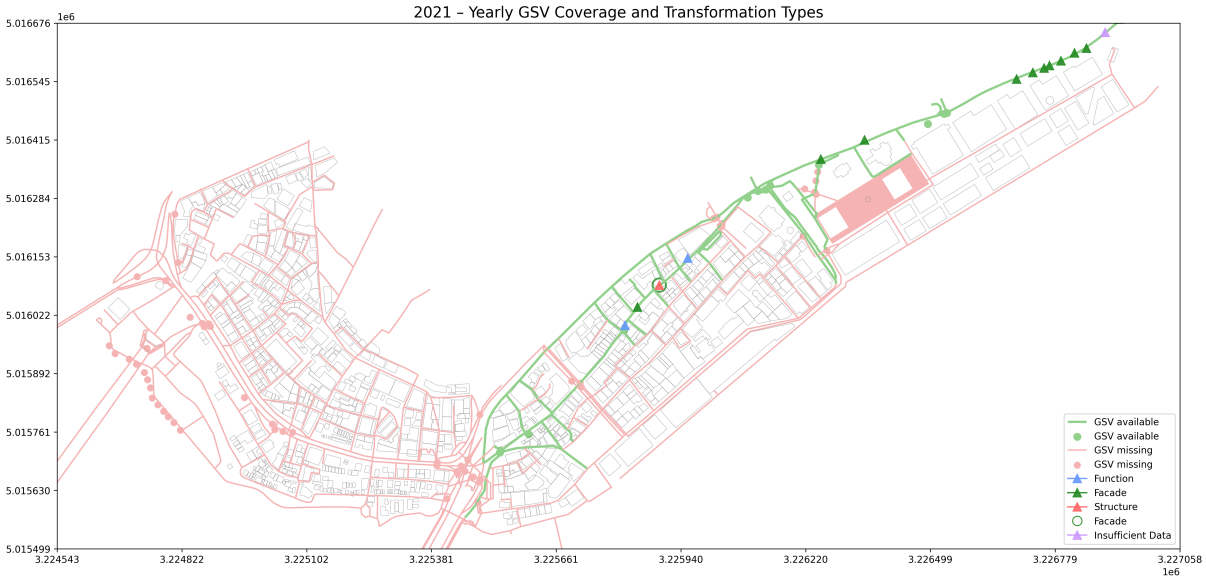


Figure 4.28: Yearly GSV Coverage and Transformation Types in 2021.

of structure and function transformations appear in the inland segments, yet they do not form any discernible clusters and remain isolated.

In 2022, spatial coverage revealed a pronounced contrast between the eastern and western sectors of the study area (Figure 4.29). On the eastern side, imagery was primarily confined to a single linear axis along Meclis-i Mebusan Street, with only limited extensions into adjacent segments such as Kılıç Ali Paşa Street. This constrained coverage resulted in sparse and narrowly distributed transformation detections, the majority of which were structural in nature. The inner streets on the eastern side remained largely undocumented, limiting the possibility of detailed transformation analysis.

In contrast, the western side displayed broader and more frequent imagery, though the coverage remained fragmented. The imagery spanned both primary routes and side streets, enabling the identification of a more varied set of transformation types. Notably, facade and function-related changes were recurrent along arterial corridors and within commercial subzones, indicating active cycles of use-based reprogramming and surface-level interventions. Although image continuity was not complete, the available data in the western section were sufficient to support localized clustering and typological differentiation, offering a more detailed spatial reading compared to the eastern side.

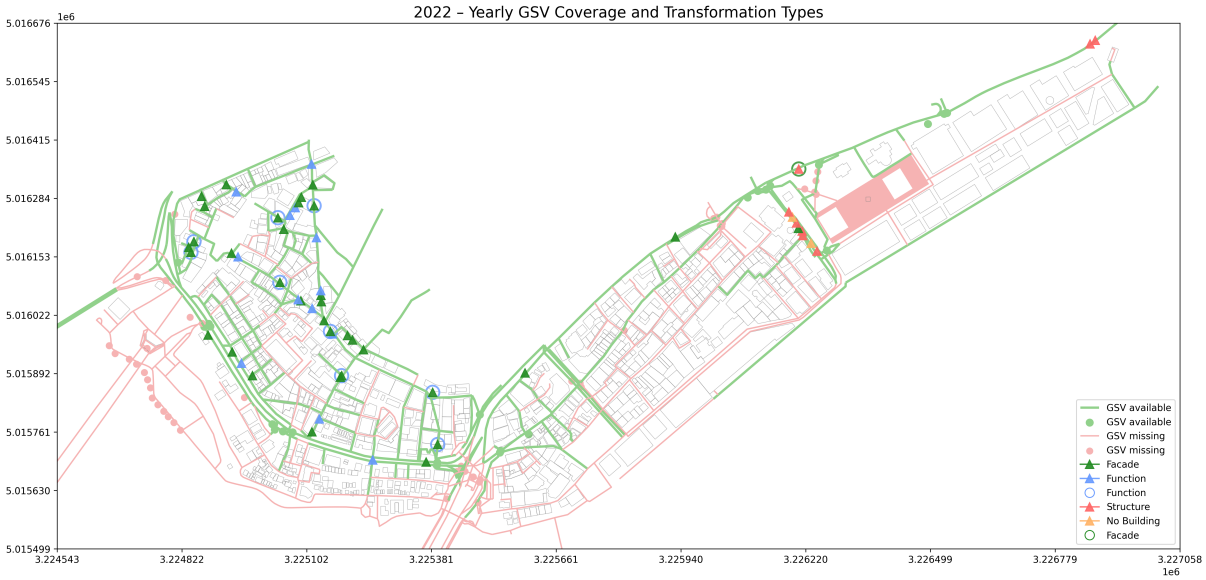


Figure 4.29: Yearly GSV Coverage and Transformation Types in 2022.

The year 2024 constitutes the most extensive and methodologically consistent phase of Street View data acquisition within the study area, featuring the highest number of georeferenced images and near-continuous spatial coverage (Figure 4.30). In contrast to earlier years characterized by fragmented or uneven imagery, the 2024 dataset demonstrates systematic and sequential capture, extending from major thoroughfares—such as Meclis-i Mebusan Street, Rıhtım Street, and Tophane İskele Street—into a broad array of secondary and tertiary streets in both eastern and western sections of the district. This level of coverage enables a more granular and spatially distributed detection of transformation signals across varied urban conditions.

Structural transformation detections are primarily concentrated in areas undergoing major redevelopment or demolition. Facade-related changes represent the most frequently identified transformation type, often appearing as linear clusters along main arteries and interior street networks. Functional changes are typically located in mixed-use zones, particularly where alterations in signage or observable programmatic shifts are evident. The frequency of combined transformations—particularly those involving both facade and function (Figure 4.31, or structure and facade—is highest in 2024, reflecting overlapping patterns of physical and use-based modification.

Notably, 2024 also functions as the primary temporal reference point in many of the

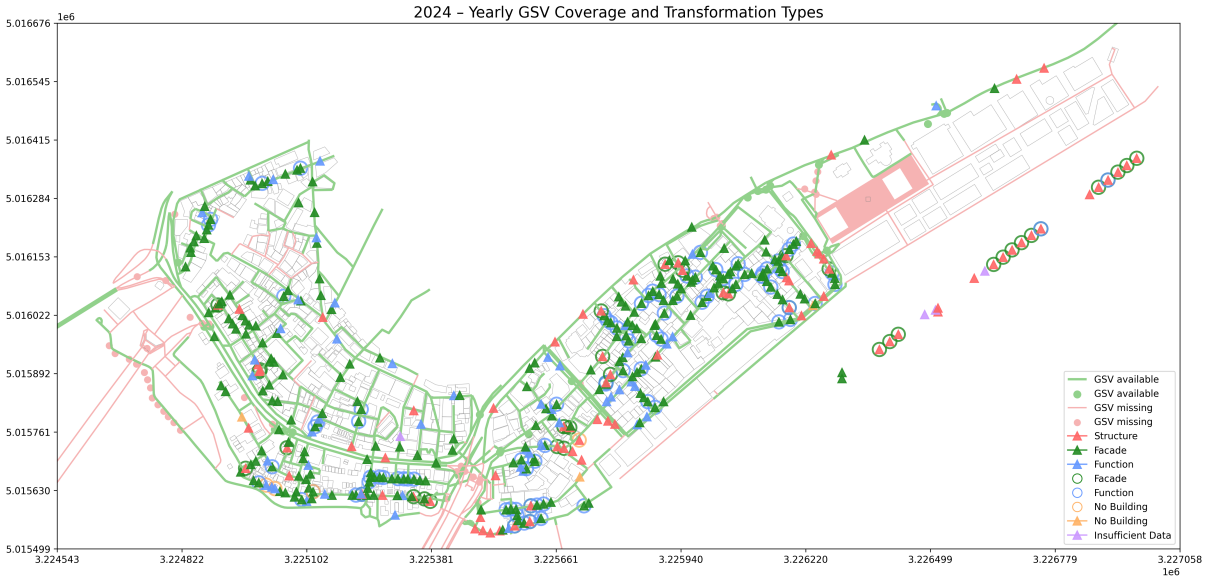


Figure 4.30: Yearly GSV Coverage and Transformation Types in 2024.



Figure 4.31: Facade transformation of a building between 2017 and 2024, with a major visible change occurring in 2024 through a pink exterior renovation.

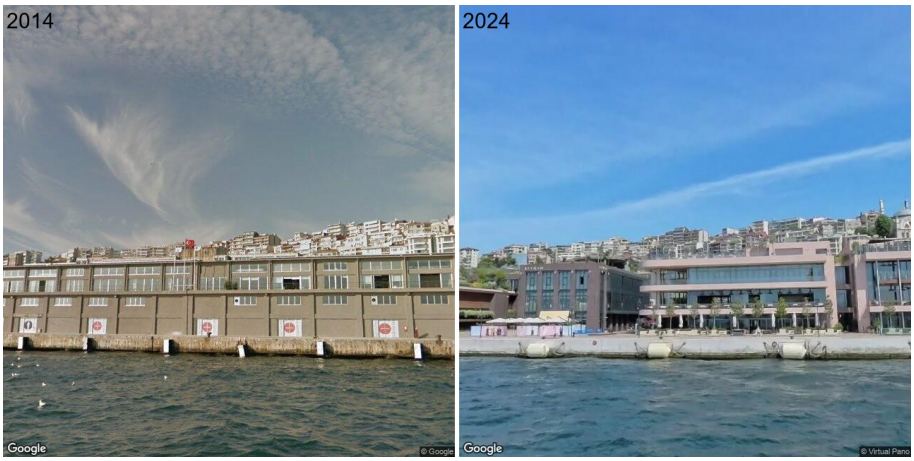


Figure 4.32: Galataport shoreline as seen from the sea, comparing views from 2014 and 2024.

model's paired-image comparisons, especially in collage-based outputs produced by the visual inference component. A substantial portion of transformation detections is derived from comparisons involving 2024 as the anchor year, which reinforces its role as a temporal benchmark. However, this also introduces a degree of temporal imbalance, as earlier years are not equally represented in comparison pairs. As a result, the elevated number of detections in 2024 partially reflects its overrepresentation in dyadic sampling, rather than a uniform temporal distribution of transformation activity. This limitation should be acknowledged in the interpretation of longitudinal patterns. For instance, one consequence of this imbalance is that certain major transformation processes—such as the Galataport redevelopment—appear to culminate in or near 2024 within the model outputs, even though construction was in fact completed significantly earlier in Figure 4.32. In such cases, the model fails to register the moment of completion due to the absence of comparable imagery from intermediate years. Indeed, for several areas along the shoreline, the only available image captured from a consistent heading after 2014 is from 2024, rendering it the sole valid reference frame for post-construction detection. This temporal gap in available Google Street View data restricts the granularity of historical interpretation and delays the visible registration of completed transformations in the system.

While these initial observations are derived from visual inspection of image sequences and the spatial distribution of transformation pings, they remain dependent on interpretive assessments and localized visual patterns. To assess whether these patterns reflect broader spatial tendencies and to reduce subjectivity in their identification, an unsupervised clustering method

based on AI is applied. The clustering process is implemented using numerically encoded features, including geographic coordinates (longitude and latitude), transformation types, and the years in which these transformations were detected. Rather than relying on visual proximity or surface-level spatial continuity, this approach groups observations based on feature similarity in a multidimensional space. As such, it enables a computational assessment of spatial regularities in transformation activity, without assuming their presence a priori.

Based on this feature-driven clustering process, the resulting spatial groupings provide a means to examine how transformation activity is distributed across the study area. The following analysis focuses on the patterns identified through the HDBSCAN algorithm. The HDBSCAN-based clustering output reveals distinct spatial patterns in the distribution of urban transformations in the Karaköy in Figure 4.33.[7] Notably, Cluster 2 (visualized in purple) is characterized by an average transformation year around 2023, predominantly associated with facade changes. However, it should be noted that GSV imagery from 2023 is not available, so this average likely reflects earlier or interpolated transformation activity. The western part of the study area exhibits relatively homogeneous patterns, with mixed instances of facade and function changes, as well as many locations with no transformation. In contrast, the southwestern section, particularly around Fermeneciler and Makaracılar streets, shows stronger signals of structural transformations. This area, known for its older and less formally documented building stock, seems to have undergone more substantial physical changes. Additionally, the waterfront zone features notable facade and function changes, albeit in a more spatially consistent manner, with only a single outlier detected in this segment.

Toward the eastern part of the area, especially near the Ziraat Bank building, clear structural transformations are observed. The presence of long-term fencing or visual obstructions in this segment may have contributed to the classification of these changes. Points 19 and 27 are particularly flagged for structural transformations, and their deviation from the main cluster results in their identification as outliers. These points likely capture overlapping structure, facade,

[7]The Appendix includes: (1) a detailed table listing the transformation years of each outlier sample, (2) illustrative examples of the extracted Google Street View images used in the study, (3) visualizations and performance scores evaluating the accuracy of the HDBSCAN clustering results, and (4) selected outcomes from human evaluation conducted to assess the coherence and validity of the clustering output.

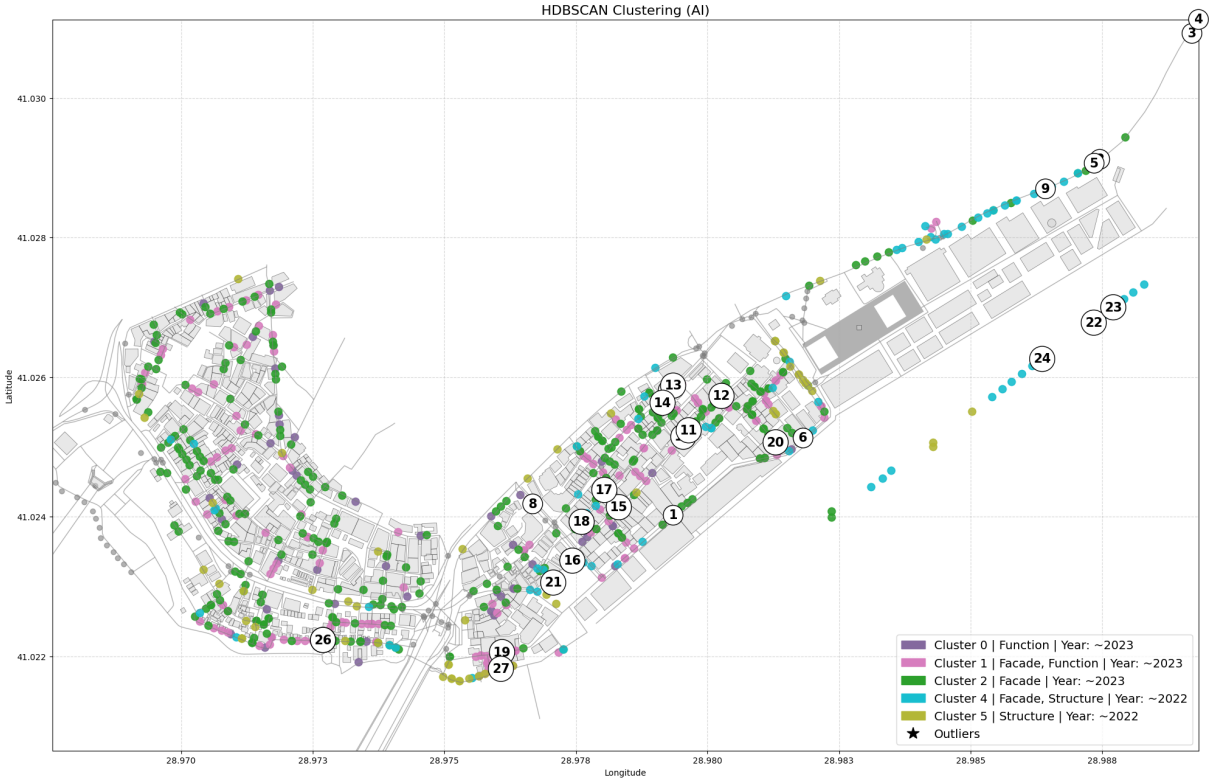


Figure 4.33: Clusters and noise points (outliers) of detected urban transformations in the Galataport area using HDBSCAN

and function changes, leading to their classification as noisy samples. Further east, a higher concentration of facade and structure transformations is detected, with these changes often documented across different construction phases. While structural changes tend to be detected in discrete events (e.g., from excavation to completion), facade and function transformations unfold more gradually, allowing more visual continuity in the dataset.

Many of the outlier samples in the east—particularly on Mumhane, Necatibey, and Kemankeş streets—represent large-scale parcel-based transformations, often related to hotel conversions or vertical extensions. Among these, notable examples include the construction of Novotel and the J.W. Marriott Hotel, as well as the demolition of the Galataport parking structure in 2024. These transformations, while significant in scale, are spatially more isolated and thus appear as noise in the clustering output. Although the Galataport zone itself might be expected to emerge as an outlier due to its overall scale and transformation intensity, its internal consistency and dense GSV sampling allow it to form its own coherent cluster. In

fact, if the spatial extent of the study were broader or the database less densely sampled in that zone, Galataport might have been classified as a noise region due to its magnitude and singularity. However, within the scale and granularity of the current dataset, the concentration of transformation points is sufficient for Galataport to emerge as a distinct cluster. That said, intermittent outlier points within Galataport still appear, most likely due to localized facade or functional transitions. Lastly, in the Tophane area, only minor changes are observed—typically small-scale facade interventions involving temporary renovation coverings (e.g., mesh or fencing)—reflecting preservation-oriented redevelopment practices.

The distribution and typology of structural, façade, and functional transformations indicate patterns of physical intervention that are both spatially concentrated and recurrent. These patterns may be associated with localized practical requirements as well as broader tendencies in the spatial organization of the urban environment. Rather than implying intentional design logics or socio-economic context, the clustering of transformation types reflects observable configurations that can be empirically identified across the study area.

Among the various segments of the study area, the eastern part of Karaköy stands out due to the density and visibility of transformation signals. This concentration aligns with literature reviews that often privilege the eastern section in discussions of urban change, largely due to its rich architectural layering and accelerated redevelopment dynamics. Accordingly, the analysis can begin by focusing on this sub-region, where empirical findings are both more numerous and more structurally complex.

A portion of the built environment within the study area holds historical value, as evidenced in the Historical Value Analysis map on the eastern side of Karaköy. According to this map, structures marked in yellow represent approximately 94% of the total building stock and are categorized as cultural heritage assets (Okyay & Binan, 2017) in Figure 4.34. This observation suggests that transformations detected by the model—particularly those classified as “structure” or “facade” changes—are not limited to physical interventions but also reflect cultural changes in the built environment.



Figure 4.34: Historical Value Distribution in the Study Area (Okyay & Binan, 2017).

In the eastern section of the study area, a notable concentration of hotel development projects is observed, often involving the restoration and adaptive reuse of both Modern Republican-era buildings and late 19th-century Ottoman structures. These interventions frequently aim to rebrand existing architectural heritage within a touristic or commercial identity framework, a tendency that aligns with broader patterns of urban commodification. In the clustering output, several outlier and noisy data points correspond to these sites of transformation, suggesting that such interventions may act as catalysts for broader shifts in spatial identity. While these points do not conform to dominant transformation patterns, their alignment with zones undergoing intensive reprogramming indicates that localized architectural heritage is being reconfigured through processes of selective preservation, commercialization, and function-driven alteration.

While the present study examines both the eastern and western sections of Karaköy using data-driven, temporally comparative methods, existing scholarship has largely concentrated on the eastern part of the neighborhood. This emphasis is often attributable to the higher density of historically registered buildings and the more apparent visibility of spatial transformation processes in that area. However, such studies frequently adopt conceptual or observational approaches rather than relying on systematically collected empirical data. In this regard, Kafescioğlu's 2020 work constitutes a notable exception, as it incorporates both the eastern and

western parts of the neighborhood within its analytical scope(Kafesçioğlu, 2016). Her use of dérive-based walking interviews offers a qualitative spatial framework that captures subjective perceptions of change across multiple subareas. While her approach is rooted in embodied spatial narratives, the current study extends this perspective by introducing a computational layer that enables the systematic detection and spatial clustering of transformation types through structured datasets.

A point of analytical convergence can be observed in areas such as Mumhane and Necati Bey Streets, where both studies independently identify intensified transformation activity. In Kafescioğlu's research, these locations are described by participants as possessing high symbolic value, often associated with notions of authenticity, selectivity, or intentional curation(Kafesçioğlu, 2016). In the present study, these same areas correspond to clusters with high frequencies of detected transformation signals, particularly in the categories of facade and function changes. This suggests a partial correspondence between locally articulated spatial perceptions and measurable patterns of physical intervention. Additionally, the presence of structural transformation signals in these zones points to the material dimension of ongoing change, indicating that such processes extend beyond surface-level modifications.

This spatial differentiation is also reflected in temporal patterns. Comparative analysis across multiple years indicates an increased frequency and intensity of transformation activity after 2020, particularly along the eastern waterfront and within the Necati Bey corridor. While Kafescioğlu's study captures Karaköy during a phase characterized by symbolic negotiation and the early stages of urban redefinition(Kafesçioğlu, 2016), the data presented here point to a more stabilized redevelopment pattern by 2025. In these locations, the spatial coincidence of symbolic value and physical intervention becomes more pronounced, with transformation signals concentrated in areas already associated with high representational and commercial significance.

By contrast, the western parts of Karaköy show relatively fewer transformation signals,

suggesting a degree of temporal and morphological continuity. However, this trend is not consistent across the entire western section. In particular, Rıhtım Street emerges as a notable exception, exhibiting high levels of transformation activity despite earlier challenges in image-based data acquisition. This indicates that transformation processes may be sensitive to local spatial conditions such as waterfront adjacency or infrastructural access, even in areas generally characterized by stability.

In aggregate, the findings suggest a partial alignment between areas associated with symbolic value—expressed through notions such as authenticity, exclusivity, or curated spatial identity—and the physical manifestations of transformation. While Kafescioğlu's study emphasizes the performative and narrative dimensions of spatial identity(Kafesçioğlu, 2016), the current analysis identifies measurable changes in the built environment that correspond with those narratively articulated zones. This points to an increasing degree of integration between symbolic framings and material outcomes, observable in the morphological adaptation of selected parts of the neighborhood.

Within this framework, facade modifications—though frequently framed as technical or aesthetic upgrades—can be interpreted as indicators of broader representational shifts. Their spatial clustering in specific corridors suggests that such interventions may play a role in repositioning the area within a new urban imaginary. These changes, when examined alongside signage strategies and commercial programming, correspond to selective alterations in the visual and functional profile of the area. Rather than merely reflecting underlying socioeconomic dynamics, these interventions appear to participate in the articulation of new spatial narratives through the built environment.

The present study brings together perceptual accounts and data-driven analyses of transformation in Karaköy, positioning them within a semiotic framework informed by Roland Barthes's theory of myth. In contrast to earlier scholarship that primarily approached spatial change through narrative and historical interpretation, the current analysis treats façade, structural, and functional alterations as empirically detectable variables. In this context, such alterations are examined not only as physical modifications but also as signifying elements that may

be associated with broader social, economic, and cultural dynamics.

When considered in relation to wider socio-spatial processes, these elements can accumulate symbolic associations that extend beyond their material form. In several cases, interventions such as façade refurbishments or adaptive reuse coincide with spatial concentrations of change that are associated with shifts in land use and visibility. While these changes may appear routine or technical, they are often situated within broader frameworks of socio-economic restructuring and urban revaluation. As such, the enhanced visibility of certain interventions may function both as indicators and potential instruments of neighborhood reconfiguration.

At the same time, the meanings ascribed to these material changes are not homogeneous. Field observations indicate the presence of multiple, and at times conflicting, interpretive frameworks—articulated through signage, branding practices, and storefront treatments—that seek to recontextualize built forms. These competing semiotic strategies appear to reflect divergent value systems, including tensions between heritage conservation and commercial rebranding, or between claims of authenticity and processes of commodification. Accordingly, Karaköy's ongoing transformation may be understood as a dynamic site of symbolic negotiation, in which material interventions and representational strategies coexist and interact across multiple registers.

Taken together, these findings point to two broader considerations. First, the spatial and temporal distribution of transformation signals suggests that material change cannot be fully disentangled from its representational framing; the built environment operates not only as a substrate for social processes but also as a medium through which they are made legible. Second, the heterogeneity in interpretive responses reinforces the importance of combining computational techniques with qualitative fieldwork. Such an integrative approach enables a more comprehensive understanding of both the observable patterns of physical change and the multiple layers of meaning that urban actors attribute to those patterns.

4.3.2 Evaluation Metrics and Model Behavior

The evaluation results characterize the model's performance in identifying multiple types of urban transformations using GSV imagery. Since the task involves a multi-label structure, where structural, functional, and facade-related changes can appear within the same image, analyzing per-class metrics and inter-class interactions is essential for quantifying detection accuracy and understanding model behavior across different transformation categories.

Overall, the micro-average performance (Precision = 0.74, Recall = 0.89, F1-score = 0.81) reveals that the model is particularly effective at retrieving relevant transformation labels across the dataset in Table 4.2. The high recall indicates that the system is highly sensitive to potential signals of change, successfully identifying most transformation occurrences. However, the lower precision suggests a tendency toward over-detection, resulting in false positives where the model detects a transformation that may not exist. This performance profile positions the model as a recall-oriented mechanism—useful for sensitive detection but in need of refinement when it comes to precision.[8]

Table 4.2: Accuracy of transformation Classifications

Class	**Precision**	**Recall**	**F1-score**	
Structure	0.73	0.80	0.76	
Facade	0.72	0.94	0.82	
Function	0.76	0.94	0.84	
No Building	0.94	0.79	0.86	
Insufficient Data	0.75	0.30	0.43	
Micro Avg	0.74	0.89	0.81	862
Macro Avg	0.78	0.75	0.74	862
Weighted Avg	0.75	0.89	0.81	862
Samples Avg	0.51	0.52	0.51	862

The macro-average metrics (Precision = 0.78, Recall = 0.75, F1-score = 0.74) show that the model maintains a relatively balanced performance across different classes, treating them

[8] **Precision**: The ratio of correctly predicted positive samples to all predicted positives. **Recall**: The ratio of correctly predicted positives to all actual positives. **F1-score**: The harmonic mean of precision and recall, indicating the balance between them. **Support**: The number of actual occurrences of each class in the dataset. **Micro Average**: Computes metrics globally by counting the total true positives, false negatives, and false positives. **Macro Average**: Calculates the metric independently for each class and then takes the unweighted mean. **Weighted Average**: Like macro average, but each class's metric is weighted by its support. **Samples Average**: Used in multi-label settings, averaging the metric over all individual samples.

equally regardless of frequency. ”Nevertheless, minority classes such as Insufficient Data and No Building slightly reduce the macro-average due to their lower detectability and inherent ambiguity. In contrast, the sample-based average F1-score is considerably lower (0.51), as it evaluates performance on a per-sample basis, which is often more challenging in multi-label classification tasks where predicting the exact set of relevant labels is difficult.”Since each image can have multiple valid labels, failing to predict any of them penalizes the model's score. This metric reveals the challenge of capturing all relevant transformation types within complex visual contexts and underscores the limitations of single-pass predictions.

Class-wise analysis provides further insight. Facade transformations exhibit one of the highest performances, with an F1-score of 0.82 and a recall of 0.94. These changes—such as new signage, repainting, or facade restructuring—are often visually salient and central within the GSV frame, making them easier for the model to detect consistently. Function-related transformations perform similarly well (F1 = 0.84, Recall = 0.94), suggesting the model effectively interprets semantically rich visual signals, particularly textual content like shop names or service icons. The slightly lower precision for this class may result from ambiguous signage or temporary elements that resemble functional transitions.

Structural changes, while still captured reasonably well (F1 = 0.76, Precision = 0.73, Recall = 0.80), tend to be more difficult to detect. These modifications—such as debris or floor additions—may occur in visually peripheral areas or be obscured by occlusions, resulting in underrepresentation or misclassification. No Building instances show the highest precision (0.94), confirming that the model rarely mislabels developed sites as empty. However, its recall (0.79) remains modest, reflecting occasional failures to detect truly vacant or undeveloped spaces—perhaps due to distant visibility, vegetation cover, or complex spatial layouts.

By contrast, Insufficient Data is the weakest-performing class, with a recall of only 0.30 and F1-score of 0.43. This category encompasses scenes with low visibility, poor lighting, or heavy occlusion, making them inherently difficult to classify. The model's struggle with this class reflects both the heterogeneous nature of such images and their underrepresentation in the training data. Improving performance here will require more targeted data collection and

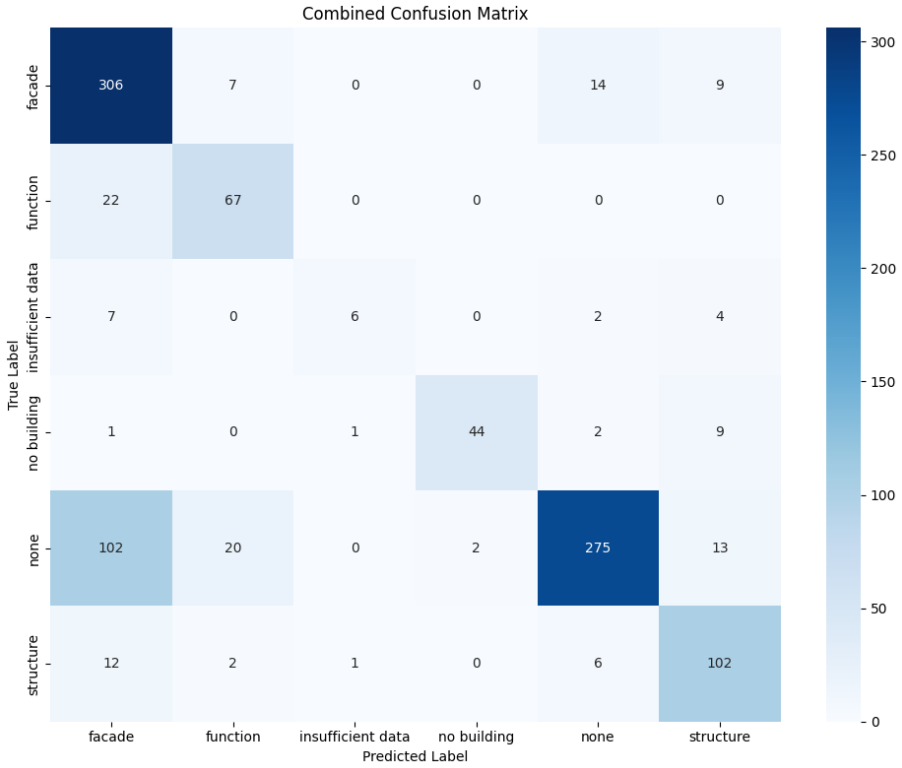


Figure 4.35: Confusion Matrix of Predicted and Ground Truth

annotation of visually ambiguous cases.

The confusion matrix reveals key patterns in the model's misclassifications in Figure 4.35. A notable number of transformation-free images ('None') are misclassified as facade changes (102 cases), artificially inflating the recall of the Facade class and reducing its precision. This suggests the model occasionally mistakes subtle visual textures or non-transformative variations for meaningful change. On the other hand, some genuine facade transformations (14 cases) are overlooked, likely because their changes are minimal or localized, and thus less visually prominent. Additional confusion exists between Function and Facade, where 22 function-related examples are labeled as facade modifications. This likely results from semantic overlap and the model's prioritization of the more visually dominant cue. A similar trend is seen in 12 cases where structural changes are misinterpreted as facade alterations, reflecting the spatial and temporal entanglement of these transformation types in real-world scenarios.

The Insufficient Data class, in particular, shows diffuse misclassifications across Facade,

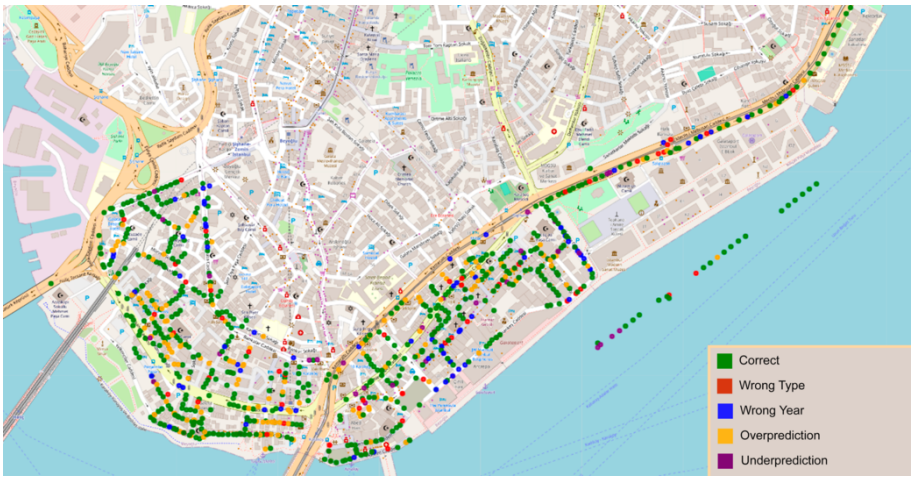


Figure 4.36: Spatial Distribution of Model Prediction Outcomes in Karaköy.

None, and Structure, indicating that such scenes lack stable visual characteristics for consistent interpretation. These cases collectively illustrate the limits of surface-level visual pattern recognition and the necessity for higher-level contextual reasoning.

Furthermore, the model's performance profile—high recall paired with moderate precision—carries epistemological implications. Rather than serving *as a definitive classifier*, the model acts as a *first-pass detection tool*, flagging scenes that may merit closer inspection. This design aligns with critical mapping and urban analysis workflows, where computational outputs are contextualized through expert review and additional sources such as zoning regulations, site histories, and on-the-ground knowledge. In this sense, the model facilitates *visual alerting* rather than final judgment, reinforcing the hybrid nature of machine-assisted urban inquiry.

The spatial distribution of prediction outcomes reveals several systematic tendencies in Figure 4.36. A high concentration of overpredictions is observed along inner streets, particularly in narrow corridors where façade changes appear exaggerated by a limited field of view. This pattern aligns with observations from human verification sessions, where the model often responded to minor alterations or partial views—changes that do not necessarily constitute full-scale façade transformations. These cases suggest a combined limitation stemming from both the constrained visual input of Google Street View and the model's sensitivity to localized architectural variations.

Instances of wrong type and wrong year predictions are scattered more evenly, often reflecting ambiguities in distinguishing overlapping transformation categories or incorrect temporal alignment. Conversely, underpredictions tend to cluster along wider roads and the Galataport shoreline, where building façades are frequently captured from oblique or distant angles. These spatial conditions likely obscure key transformation features, leading the model to miss relevant change signals. Overall, the map illustrates how prediction errors are spatially patterned and context-dependent, shaped by both the geometry of the street network and the visual framing limitations inherent in the data source.

From a theoretical and methodological perspective, these results reinforce one of the thesis's core arguments: that urban visual transformations are semantically entangled, context-dependent, and computationally ambiguous. The frequent confusion between facade and structure, or the model's occasional detection of changes in static environments, highlights the limitations of pixel- and patch-level classification when used in isolation. These findings validate the use of Visual Language Models (VLMs) and the integration of Chain-of-Thought (CoT) prompting strategies, which introduce interpretive layers that help the model to disambiguate overlapping signals and ground predictions in semantically coherent outputs.

Chapter 5
Conclusion

5.1 General Evaluation

This study establishes a methodological framework that employs Google Street View imagery, in conjunction with vision-language models and diachronic visual analysis techniques, to identify and spatially contextualize physical changes in the urban environment. The proposed approach integrates temporal sequencing, panoramic reconstruction, and automated captioning to detect transformations in façade, function, and structure. These changes, once georeferenced and clustered, facilitate the observation of spatial and temporal patterns across the study area. While the model effectively identifies transformation signals and produces measurable outputs in specific zones, such as Mumhane and Necati Bey Streets, its function remains predominantly detection-oriented. Quantitative results confirm that the pipeline is capable of supporting large-scale visual segmentation tasks; however, the interpretive depth of the findings is constrained by the absence of localized semantic enrichment and post-hoc qualitative verification.

Several methodological limitations affect the generalizability and precision of the system. The resolution and consistency of Street View imagery, variations in pitch and field of view, and the lack of fine-tuning on domain-specific architectural datasets contribute to misclassification in visually ambiguous scenes. Moreover, the foundation model employed—based on GPT-4.1—operates as a closed system, limiting transparency and interpretability. While additional models such as Gemma 3, AyaVision, and LLaVA were evaluated in preliminary tests, none demonstrated superior performance in this domain without further customization. Despite these constraints, the study contributes to the development of computational tools for the systematic

documentation of urban change. The resulting framework provides a scalable mechanism for transformation detection and a structured basis for further spatial analysis, offering potential utility in urban research, policy monitoring, and digital archiving.

In addition to these methodological constraints, when examined through the lens of critical cartography, one of the key observations of this study is the structural inconsistency and epistemic limitations of Google Street View (GSV) as a primary data source for analyzing urban transformation. Contrary to prevailing assumptions about the comprehensiveness of GSV's coverage, the data obtained through this platform is characterized by significant spatial and temporal gaps. In particular, images captured during specific years, such as 2021, display sparse or incomplete coverage for numerous locations. This lack of coverage appears to stem from logistical disruptions caused by the COVID-19 pandemic, combined with the widespread closure of public spaces and a broader slowdown in socioeconomic activity, which likely reduced the perceived relevance of documenting urban change during that period. As a result, critical moments such as the pandemic are largely absent from GSV imagery, even though the pandemic had a profound impact on city life. This absence underscores the selective and goal-oriented nature of map-making, raising important questions about through institutional choices and intended purposes. Some streets have entirely missing data, while others present only partial visual records with restricted viewing angles. These discrepancies are neither random nor merely technical limitations; rather, they reflect broader operational logics determined by commercial priorities, infrastructural constraints, and algorithmic decisions inherent to the platform. Consequently, areas lacking economic appeal or visibility remain systematically unrecorded, effectively removing them from analytical visibility and scholarly scrutiny. From a critical cartography perspective, such selective visibility underscores the power dynamics and ideological frameworks embedded within digital cartographic practices, as platforms like GSV exercise significant authority over spatial representation and urban epistemology.

Furthermore, visual inconsistencies within the available GSV images further challenge reliable comparative analysis, aligning closely with critical cartography's emphasis on the constructed and contingent nature of maps. Variations due to environmental conditions, occlusions, inconsistent lighting, or shifts in camera positioning result in substantial disruptions to visual

continuity, complicating objective assessments of physical change. Such discrepancies often cause computational models, particularly vision-language models (VLMs), to misinterpret these visual variations as actual transformations. As these models rely heavily on pixel-level consistency and feature alignment, inconsistencies introduce considerable interpretative uncertainty. Consequently, identified transformations may represent technological artifacts rather than empirical urban changes, highlighting how visual representations are mediated and potentially misleading.

The integration of AI into this methodological framework introduces additional layers of interpretative complexity, resonating with critical cartography's critique of epistemological neutrality. These models interpret visual data based on linguistic prompts, which inherently introduce a layer of discursive subjectivity. The wording and structure of these prompts significantly shape the models' outputs, making their interpretations contingent upon the language used. For example, subtle differences in phrasing can dramatically alter the detection and classification of functional transformations, demonstrating the inherently constructed nature of analytical outcomes. Moreover, aligning these outputs with general typologies or dominant representations further reinforces specific normative frameworks. This reliance on linguistic framing positions the maps produced through this methodology as discursive artifacts, shaped by infrastructural biases, computational constraints, and epistemic assumptions. Recognizing these dynamics highlights the critical cartography principle that mapping is an inherently political act, necessitating ongoing scrutiny of the assumptions and limitations embedded in cartographic methods and outcomes.

5.2 Limitations of the method

This study could make contributions at the intersection of computer vision, mapping theory, and urban transformation; nevertheless, it is bounded by a set of interlocking limitations that shape the reliability, interpretability, and generalisability of its findings. The Google Street View imagery used in this study is characterized by temporal and spatial inconsistency. While certain streets are captured multiple times within a given year, others exhibit multi-year gaps.

For example, the western section of the Galataport area lacks image coverage for the year 2021, limiting the availability of continuous visual data for longitudinal analysis. Because such gaps interrupt the diachronic sequence, they compromise our ability to trace continuous transformation trajectories. Even where imagery is available, exogenous capture conditions such as varying sun angle, seasonal foliage, moving vehicles, or temporary hoardings introduce visual noise that the model can misread as genuine change. An additional limitation arises from the visual artifacts introduced during the creation of stitched panoramas. In some collage images, particularly those covering wide temporal ranges (e.g., 2012–2020), text labels or visual elements embedded in the image layout (such as year markers in similar colors) may become difficult to distinguish.

From a methodological perspective, the taxonomy of transformation types—facade, function, and structure—remains inherently ambiguous in certain cases. Visual cues such as signage alterations can vary significantly in semantic weight: while some signage changes reflect a genuine functional shift or substantial facade redesign, others constitute only minor visual updates without broader architectural or commercial implications. The model, however, may interpret any visible alteration in signage as a facade transformation, even when such changes fall outside the conventional definition of a facade intervention. This is partly due to the model's sensitivity to surface-level pixel changes and its learned tendency to associate visible variation with categorical transformation. In some cases, despite explicit prompts indicating functional change, the model may instead prioritize aesthetic signals and classify the scene as a facade modification. A related issue arises from temporally contingent elements such as metal shutters. These shutters may appear closed in one frame and open in another due to time-of-day differences in image capture. While no structural or functional change has occurred, the model often classifies such variations as facade transformations, especially when the closed shutter occupies a large visual area. This classification persists even when explicitly addressed in the prompt. Furthermore, if the visual language model (VLM) fails to correctly describe the shutter in its output—for instance, by omitting the term altogether or describing it as something else-LLM may assign disproportionate interpretive weight to the visual difference, resulting in misclassification

Moreover, the model demonstrates a consistent tendency to overreact to minor visual variations, particularly in narrow urban streets. This oversensitivity appears to result from both the limitations of the input data and the model's interpretation strategy in Figure 4.36. In confined spatial settings, Google Street View imagery often captures only partial views of buildings and streetscapes due to occlusions, limited depth of field, or restricted camera angles. As a result, even minor elements—such as partially closed shutters, small signage updates, or changes in lighting—can occupy a disproportionately large area in the image. When such localized changes dominate the visible frame, the model is more likely to interpret them as meaningful transformations, even though they may lack architectural or functional significance in reality. This perceptual amplification is further exacerbated by the model's lack of contextual grounding, leading to classification outputs that may not reflect the actual extent or type of change. Many of the data inconsistencies and false positives observed in this study can be traced back to this combination of constrained visual input and over-responsive interpretation, reflecting a compounded limitation stemming from both the data source (GSV) and the model itself.

The foundation model employed in this study—GPT-4.1—operates within a closed architecture. Its internal parameters, such as training data composition, weight initialization, and attention mechanisms, are not publicly disclosed. This opacity restricts the ability to trace the specific sources of model behaviors or classification errors. Moreover, the model has not been fine-tuned on domain-specific data related to urban transformation, which likely reduces its capacity to differentiate between visually similar yet semantically distinct changes.

While the proposed pipeline provides a robust framework for identifying visual indicators of transformation through Google Street View (GSV), its analytical scope is inevitably shaped by the affordances and constraints of the platform itself. GSV's vehicle-bound perspective, selective update cycles, and platform-imposed redactions result in a form of visual omission that is rarely addressed in change detection workflows. Although not originally designed for this purpose, platforms like Mapillary offer a highly suitable environment for counter-mapping practices and are well-positioned to be utilized in this context. These practices reveal aspects of the built environment that remain undocumented by GSV and emphasize forms of spatial

documentation that emerge from individual, localized interests.[1]

Unlike traditional cartographic representations, which operate through symbolic abstraction, GSV's image-based platform reflects a commercial logic of visibility, bounded by vehicular accessibility and platform intent. While maps abstract space from above, GSV simulates immersion from below—but both omit as much as they reveal. In this sense, the seeming neutrality of GSV may be read as a visual myth that naturalizes a selective and contingent urban gaze.

To assess comparative performance, several additional multimodal models were also tested during the exploratory phase, including Google Gemma 3, AyaVision, and LLaVA. Although each demonstrated some capability in visual reasoning, none consistently outperformed the GPT-based model in tasks specific to urban transformation. These limitations suggest that without fine-tuning on geospatial or architectural datasets, general-purpose vision-language models remain insufficient for reliably capturing the nuances of built environment change.

5.3 Future Directions and Applications

The methodological and technical limitations outlined in this study also serve to define clear avenues for future development and application. One of the primary challenges identified is the lack of domain-specific fine-tuning in the underlying foundation model, which currently operates on general-purpose visual-textual representations. Future research could address this by incorporating task-specific fine-tuning on urban imagery, particularly in relation to architectural typologies, signage conventions, and region-specific urban textures. Fine-tuning the vision-language model (VLM) on annotated datasets containing validated transformation types would likely improve classification accuracy, especially for subtle or ambiguous cases.

Another critical direction involves improving spatial and temporal alignment between image sequences. Current misclassifications—such as overinterpreting minor edits in narrow streets or failing to detect building absence—could be reduced through the integration

[1] See Appendix for a comparative example illustrating this representational gap

of high-precision geo-registration and viewpoint normalization tools. In future applications, incorporating urban morphological attributes—such as street width, building setbacks, and spatial enclosure—may allow for dynamic adjustment of image acquisition parameters. Specifically, pitch and field-of-view (FOV) settings could be calibrated during Google Street View downloads to better capture façades in morphologically constrained environments. Alternatively, a custom data collection mechanism could be deployed to systematically capture panoramic images at fixed spatial and temporal intervals. Such strategies would help address both visual distortion and the diachronic gaps that currently limit transformation detection.

The integration of spatial priors and auxiliary data sources (e.g., zoning plans, cadastral maps, or municipal redevelopment records) offers another promising direction. Multimodal models could be extended to condition visual interpretation on non-visual contextual layers, allowing for more robust interpretation of ambiguous or occluded scenes. Additionally, the development of explicit attention mechanisms or spatial grounding modules would improve the system's ability to distinguish between multiple co-present buildings, mitigating the attribution errors observed in dense urban collages. Beyond architectural integration, complementary annotation strategies could also be pursued. For instance, creating a human-verified dataset drawn from building stocks that exhibit similar morphological characteristics and transformation patterns would allow for more descriptive, semantically consistent, and interpretable classification outcomes. Such curated verification sets—especially if produced by multiple annotators—would clarify ambiguous transformation labels and strengthen the empirical basis for model evaluation.

In terms of practical applications, the proposed method has strong potential to support urban monitoring, historical mapping, and post-disaster assessment tasks. Municipalities and planning agencies could adapt the pipeline to track facade compliance, signage turnover, or informal construction in time-sensitive workflows. Likewise, heritage conservation efforts could leverage temporally annotated transformation maps to monitor alterations in historically protected districts.

The method could be applied to a range of urban environments with different morphological and socio-economic conditions. However, adaptation to new contexts would require benchmarking the pipeline against locally specific architectural features and visual practices. The general framework may also be integrated into larger geospatial platforms, enabling automated spatial annotation of visual change and supporting longitudinal studies in urban informatics, visual culture, or planning policy.

Beyond the case of Karaköy, the proposed method holds potential for application across diverse urban settings, particularly those exhibiting varying morphological and socio-economic structures. This includes sites such as Tarlabaşı in Istanbul and Ulus in Ankara, both characterized by contested renewal processes and layered historical fabrics, as well as international districts like Neukölln in Berlin and Brooklyn in New York City, where visual indicators of transformation intersect with processes of gentrification, migration, and shifting socio-economic identities. Adaptation to new contexts, however, would necessitate benchmarking the pipeline against local architectural typologies and visual signs. More broadly, the framework contributes to the construction of a structured visual archive of urban transformation, an effort that extends beyond mere detection toward the systematic recording, interpretation, and semantic organization of change over time. By treating artificial intelligence not only as a detection tool but as an interpretive agent, the approach facilitates the translation of image-based evidence into readable urban narratives.

References

Acar, Y. (2019). *Bilginin haritalanması: Bilgi, ilişkilendirme*. Pinhan Yayıncılık.

Acemoglu, D. (2021). *Redesigning ai*. MIT Press.

Adriaans, P. (2021). *Information* [The Stanford Encyclopedia of Philosophy (Fall 2021 Edition), edited by Edward N. Zalta. Accessed: 2025-06-18].

Aksu, B., & Tutar, C. (2021). Neoliberal toplumsal dinamikler ekseninde kentsel mekânların kurgusu: Karaköy örneği. *Etkileşim*, (8), 108–131.

Alcantarilla, P. F., Stent, S., Ros, G., Arroyo, R., & Gherardi, R. (2018). Street-view change detection with deconvolutional networks. *Autonomous Robots*, *42*(7), 1301–1322.

Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., & Weaver, J. (2010). Google street view: Capturing the world at street level. *Computer*, *43*(6), 32–38.

Arseven, C. E., & Yelkenci, D. (1989). *Eski galata ve binâları* (Vol. 1). Çelik Gülersoy Vakfı.

Bahçekapılı, Y. (2020). Mekânsal ayrışmayı yürüyerek okumak: Karaköy'ün hayali sınırları. *İdealkent*, *11*(Kentleşme ve Ekonomi Özel Sayısı), 1379–1406.

Barthes, R. (2015). *Mythologies*. Média Diffusion.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and gis: A review. *Landscape and Urban Planning*, *215*, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Bilsel, C. (2011). "les transformations d'istanbul": Henri prost's planning of istanbul (1936-1951). *A— Z ITU Journal of the Faculty of Architecture*, *8*(01), 100–116.

Boole, G. (1847). *The mathematical analysis of logic*. Philosophical Library.

Boonstra, L. (2025). Prompt engineering [Contributors: Michael Sherman, Yuan Cao, Erick Armbrust, Anant Nawalgaria, Antonio Gulli, Simone Cammel; Editors: Antonio Gulli,

Anant Nawalgaria, Grace Mollison]. https : / / cloud . google . com / docs / vertex - ai / generative-ai/docs/prompt-engineering

Cinnamon, J., & Gaffney, A. (2022). Do-it-yourself street views and the urban imaginary of google street view. *Journal of urban technology*, *29*(3), 95–116.

Crampton, J. W. (2011). *Mapping: A critical introduction to cartography and gis*. John Wiley & Sons.

Crampton, J. W., & Krygier, J. (2005). An introduction to critical cartography. *ACME: An International Journal for Critical Geographies*, *4*(1), 11–33.

Curtis, J. W., Curtis, A., Mapes, J., Szell, A. B., & Cinderich, A. (2013). Using google street view for systematic observation of the built environment: Analysis of spatio-temporal instability of imagery dates. *International journal of health geographics*, *12*, 1–10.

Dağdelen, İ. (2021). Osmanlı haritalarında istanbul [Accessed: June 26, 2025].

Dodge, M., & Kitchin, R. (2007). 'outlines of a world coming into existence': Pervasive computing and the ethics of forgetting. *Environment and planning B: planning and design*, *34*(3), 431–445.

Eldem, E. (2000). *Bankalar caddesi: Osmanlı'dan günümüze voyvoda caddesi*. Osmanlı Bankası Bankacılık ve Finans Tarihi Araştırma ve Belge Merkezi.

Erbil, A. Ö., & Erbil, T. (2001). Redevelopment of karaköy harbor, istanbul: Need for a new planning approach in the midst of change. *Cities*, *18*(3), 185–192.

Erkökoğlu, A. (2002). Beyoğlu planlama sürecine bakış. *Planlama*, (2).

ETH Zürich. (2025). Future cities lab (fcl) [Erişim: 2025-07-06].

Fan, Z., Zhang, F., Loo, B. P., & Ratti, C. (2023). Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences*, *120*(27), e2220417120.

Foucault, M. (2013). *Archaeology of knowledge*. routledge.

Foucault, M. (2020). Power/knowledge. In *The new social theory reader* (pp. 73–79). Routledge.

Fry, D., Mooney, S. J., Rodríguez, D. A., Caiaffa, W. T., & Lovasi, G. S. (2020). Assessing google street view image availability in latin american cities. *Journal of Urban Health*, *97*(4), 552–560. https://doi.org/10.1007/s11524-020-00442-4

Gilge, C. (2016). Google street view and the image as experience. *GeoHumanities*, *2*(2), 469–484.

Giray, L. (2023). Prompt engineering with chatgpt: A guide for academic writers. *Annals of biomedical engineering*, *51*(12), 2629–2633.

Goad, C. E. (1905). Insurance plan of constantinople: Galata and pera [Digitized by SALT Research. Accessed June 2025].

Gonzalez, D., Rueda-Plata, D., Acevedo, A. B., Duque, J. C., Ramos-Pollán, R., Betancourt, A., & García, S. (2020). Automatic detection of building typology using deep learning methods on street level images. *Building and Environment*, *177*, 106805.

Google. (n.d.). How street view works and where we will collect images next [Accessed: 2025-07-03]. https://www.google.com/streetview/how-it-works/

Graafland, A. D. (2010). *Understanding the socius through creative mapping techniques*. Delft School of Design, Delft University of Technology.

Harley, J. B. (1989). Deconstructing the map. *Cartographica: The international journal for geographic information and geovisualization*, *26*(2), 1–20.

Harley, J. B. (2018). The map and the development of the history of cartography.

Huang, T., Wu, Z., Wu, J., Hwang, J., & Rajagopal, R. (2024). Citypulse: Fine-grained assessment of urban change with street view time series. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(20), 22123–22131.

Kafesçioğlu, F. O. (2016). 19. yüzyılın ikinci yarısından günümüze galata/karaköy'de kent morfolojisi ve yapı türlerinin incelenmesi. *İdealkent*, *7*(18), 174–203.

Karaman, O. (2008). Urban pulse—(re) making space for globalization in istanbul. *Urban Geography*, *29*(6), 518–525.

Kim, J., & Jang, K. M. (2023). An examination of the spatial coverage and temporal variability of google street view (gsv) images in small-and medium-sized cities: A people-based approach. *Computers, Environment and Urban Systems*, *102*, 101956.

Klimenka, M., Zhao, K., Hilland, R., Zhang, F., Voogt, J., & Ratti, C. (2025). Instant infrared: Estimating urban surface temperatures from street view imagery. *Building and Environment*, *267*, 112122.

Kollektiv Orangotango. (2018). *This is not an atlas: A global collection of counter-cartographies* [eBook]. transcript Verlag.

Küçük, S. G., & Mazlum, D. (2017). Karaköy-kemeraltı bölgesinde 19. yüzyıldan günümüze yaşanan kentsel ve toplumsal gelişim, dönüşüm ve koruma sorunları: Karaköy-kemeraltı bölgesinde 19. yüzyıldan günümüze yaşanan kentsel ve toplumsal gelişim, dönüşüm ve koruma sorunları. *Tasarim+ Kuram*, *13*(24), 90–111.

Lefebvre, H. (2012). From the production of space. In *Theatre and performance design* (pp. 81–84). Routledge.

Leibnitz, G.-G. (1703). Explication de l'arithmétique binaire, qui se sert des seuls caractères o et i avec des remarques sur son utilité et sur ce qu'elle donne le sens des anciennes figures chinoises de fohy. *Mémoires de mathématique et de physique de l'Académie royale des sciences*.

Li, X. (2021). Examining the spatial distribution and temporal change of the green view index in new york city using google street view images and deep learning. *EPB: Urban Analytics and City Science*, *48*(7), 2039–2054. https://doi.org/10.1177/2399808320962511

Li, X., Yoshimura, Y., Tu, W., & Ratti, C. (2022). Quantifying neighborhood visual disorder using street view image time series. *Landscape and Urban Planning*, *222*, 104382. https://doi.org/10.1016/j.landurbplan.2022.104382

Liu, C., & Song, W. (2024). Mapping property redevelopment via geoai: Integrating computer vision and socioenvironmental patterns and processes. *Cities*, *144*, 104644.

Liu, Y., Wang, Z., Ren, S., Chen, R., Shen, Y., & Biljecki, F. (2025). Physical urban change and its socio-environmental impact: Insights from street view imagery. *Computers, Environment and Urban Systems*, *119*, 102284.

Lynch, K. (1964). *The image of the city*. MIT press.

McConchie, A. (2015). Hacker cartography: Crowdsourced geography, openstreetmap, and the hacker political imaginary. *ACME: An International Journal for Critical Geographies*, *14*(3), 874–898.

Merleau-Ponty, M., & Smith, C. (2015). *Phenomenology of perception* (Vol. 26). Routledge London.

MIT Senseable City Lab. (n.d.). Mit senseable city lab [Accessed: 2025-07-06].

Monmonier, M. (2018). *How to lie with maps*. University of Chicago Press.

Müller-Wiener, W. (2001). *Istanbul'un tarihsel topografyası* (Ü. Sayın, Trans.). İstanbul.

Navalgund, R. R., Jayaraman, V., & Roy, P. (2007). Remote sensing applications: An overview. *current science*, 1747–1766.

Newman, W. E. (2024). Translations from coding to reality.

NUS Urban Analytics Lab. (n.d.). Urban analytics lab [Accessed: 2025-07-06].

Ogawa, Y., Zhao, C., Oki, T., Chen, S., & Sekimoto, Y. (2023). Deep learning approach for classifying the built year and structure of individual buildings by automatically linking street view images and gis building data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 1740–1755.

Okyay, G. G., & Binan, D. U. (2017). Bir kültürel miras alanı olarak karaköy-tophane bölgesi'nin sahip olduğu değerlerin irdelenmesi: Bir kültürel miras alanı olarak karaköy-tophane bölgesi'nin sahip olduğu değerlerin irdelenmesi. *Tasarim+ Kuram*, *13*(23), 15–38.

Parkinson, H. J. (2014, October). *Google puts street view trekker camera on camel to document liwa desert* [Published: 10 Oct 2014]. https://www.theguardian.com/technology/2014/oct/10/google-camel-street-view-trekker-camera-documents-liwa-desert

Pickles, J. (2004). *A history of spaces: Cartographic reason, mapping and the geo-coded world*. Routledge.

Quinn, S., & Alvarez León, L. F. (2019). Interfacing urban intelligence: Google street view and the spatial dimensions of algorithmic bias. *Urban Geography*, *41*(7), 939–948. https://doi.org/10.1080/02723638.2019.1638354

Raghu, D., Bucher, M. J. J., & De Wolf, C. (2023). Towards a 'resource cadastre'for a circular economy–urban-scale building material detection using street view imagery and computer vision. *Resources, Conservation and Recycling*, *198*, 107140.

SALT. (2021). Voyvoda caddesi ve bankalar caddesi [SALT Araştırma Arşivi üzerinden erişildi]. https://archives.saltresearch.org/handle/123456789/114287

Sarı, T., Bayar, R., & Burak, N. (2024). From commerce to art: Transformation of intervention through the interactions of virtual and urban space in karaköy. *Journal of Design for Resilience in Architecture Planning*, *5*(3), 342–362. https://doi.org/10.47818/DRArch.2024.v5i3136

Segal, P. Z., & Acres, 5. (n.d.). Tools for community land access advocacy [Accessed: 2025-07-01]. *596 Acres*.

Şehir Planlama Müdürlüğü. (n.d.). Galataport Projesi [Erişim adresi: https://sehirplanlama.ibb.istanbul/galataport-projesi/]. https://sehirplanlama.ibb.istanbul/galataport-projesi/

Sequoiah-Grayson, S., & Floridi, L. (2022). Semantic conceptions of information (E. N. Zalta, Ed.; Spring 2022) [Accessed: 2025-06-18]. *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/spr2022/entries/information-semantic/

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423.

Stanford CityBlock Project [Accessed: 2025-06-03]. (2013).

Sun, Y., Zhang, D., Duarte, F., & Ratti, C. (2022). Using deep learning to examine street view greenery and its association with self-reported physical activity across 11 countries. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-05545-w

Szcześniak, J. T., Ang, Y. Q., Letellier-Duchesne, S., & Reinhart, C. F. (2022). A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations. *Building and Environment*, *207*, 108108.

Tepeli, Ö., & Ocakçı, M. (2017). The change of identity and memory on urban space with project impact: Karaköy kemeraltı district. *ICONARCH International Congress of Architecture and Planning; 2017: ICONARCH III-Memory of Place in Architecture and Planning*.

Vallebueno, A., & Lee, Y. S. (2023). Measuring urban quality and change through the detection of physical attributes of decay. *Scientific reports*, *13*(1), 17316.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Vega, D. C. G. d. (2016, August). *Sheep view: Where there's a wool, there's a way* [Accessed: 2025-06-18].

Wang, S., & Han, J. (2024). Automated detection of exterior cladding material in urban area from street view images using deep learning. *Journal of Building Engineering*, *96*, 110466.

Wang, Z., Ito, K., & Biljecki, F. (2024). Assessing the equity and evolution of urban visual perceptual quality with time series street view imagery. *Cities*, *145*, 104704.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

WIRED. (2022). All the ways google gets street view images [Interview with Ethan Russell, Director of Product, Google Maps. Accessed: 2025-06-03].

Wood, D. (1992). *The power of maps*. Guilford Press.

Zajc, M. (2015). Social media, prosumption, and dispositives: New mechanisms of the construction of subjectivity. *Journal of Consumer Culture*, *15*(1), 28–47. https://doi.org/10.1177/1469540513493201

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, *180*, 148–160.

Zhang, L., Pei, T., Wang, X., Wu, M., Song, C., Guo, S., & Chen, Y. (2020). Quantifying the urban visual perception of chinese traditional-style building with street view images. *applied sciences*, *10*(17), 5963.

Zhang, Y., Liu, P., & Biljecki, F. (2023). Knowledge and topology: A two layer spatially dependent graph neural networks to identify urban functions with time-series street view image. *ISPRS Journal of Photogrammetry and Remote Sensing*, *198*, 153–168.

Zhou, H., Liu, L., Lan, M., Zhu, W., Song, G., Jing, F., Zhong, Y., Su, Z., & Gu, X. (2021). Using google street view imagery to capture micro built environment characteristics in drug places, compared with street robbery. *Computers, Environment and Urban Systems*, *88*, 101631.

Zhou, H., He, S., Cai, Y., Wang, M., & Su, S. (2019). Social inequalities in neighborhood visual walkability: Using street view imagery and deep learning technologies to facilitate healthy city planning. *Sustainable cities and society*, *50*, 101605.

Appendix

Table A.1: Detected Years of Transformation in Outlier Samples

#	Structure	Facade	Function
1	2015	2019	2020
2	2022	None	None
3	None	None	None
4	None	None	None
5	2022	None	None
6	2020	2024	2024
7	2015	2018	2021
8	2018	2019	2024
9	2024	None	None
10	2015	None	2024
11	2015	2024	2024
12	2017	2024	2024
13	2015	2018	2024
14	2024	2024	2018
15	2019	2024	2024
16	2024	2020	2020
17	2018	None	2024
18	2024	None	2024
19	2024	2024	2024
20	2024	2024	2024
21	2024	2018	2018
22	2024	None	None
23	2024	2024	2024
24	2024	2024	2024
25	2024	2020	2020
26	2024	2020	2020
27	2019	2019	2024

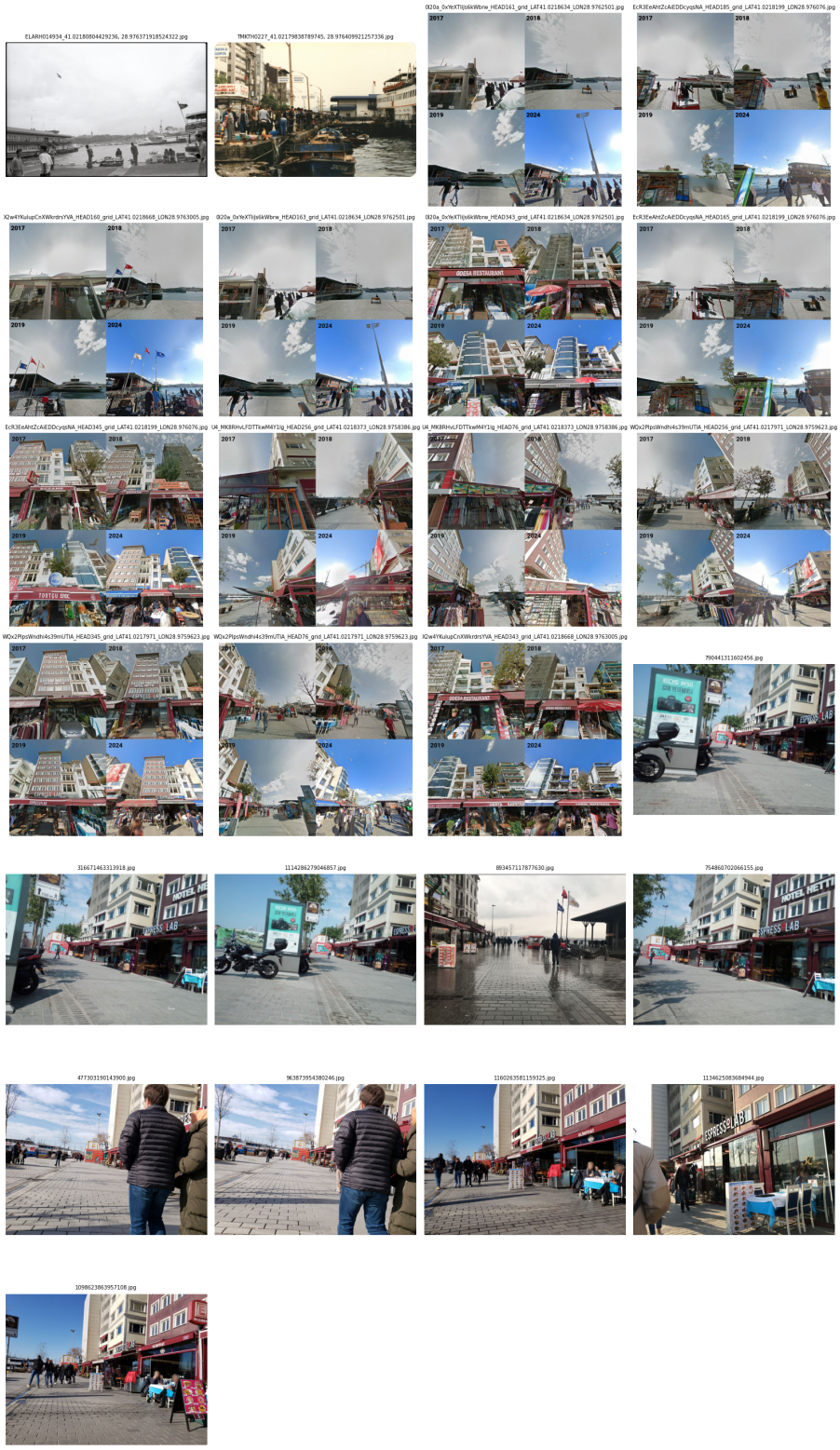


Figure A.1: Composite visual layout of archival and GSV AND Mapillary images associated with multiple urban characters across different time periods and locations.

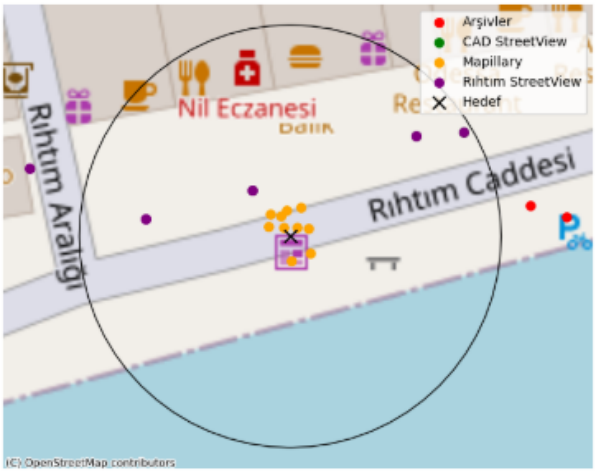


Figure A.2: OSM Map shows the location of archival images.

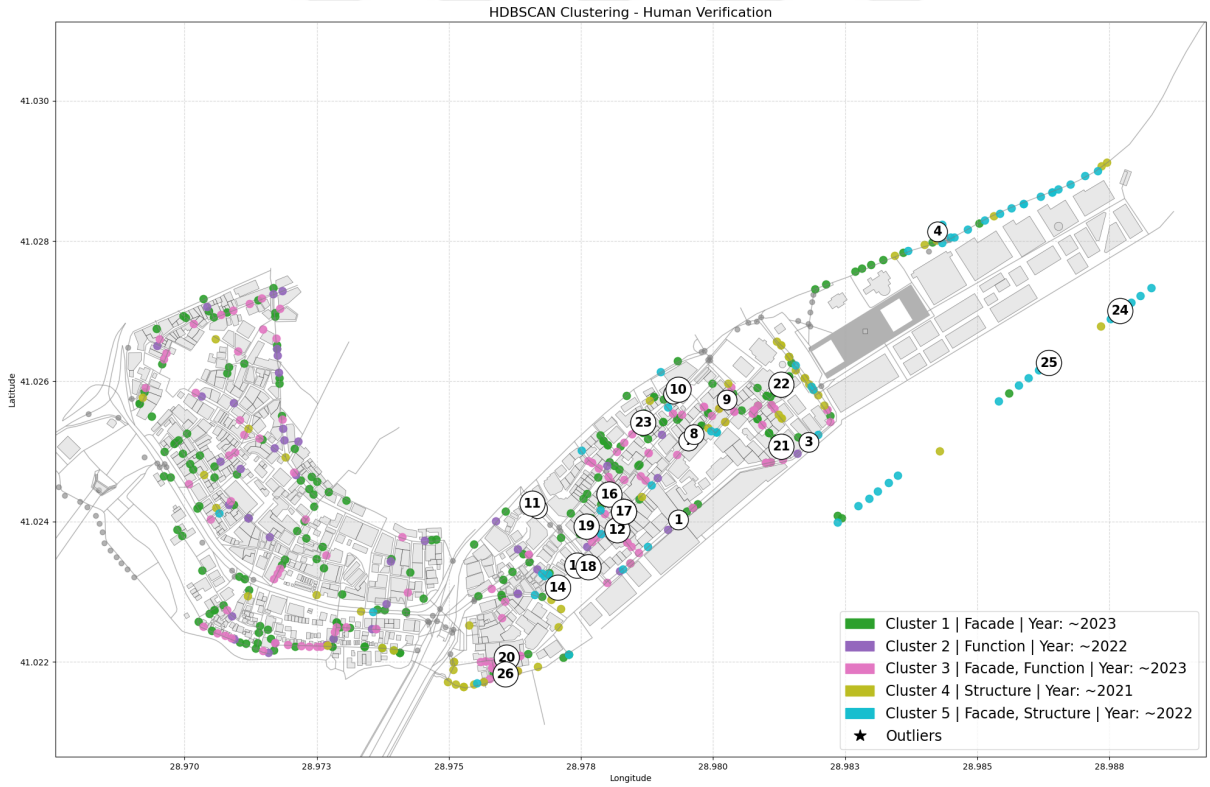


Figure A.3: Spatial distribution of transformation clusters produced by the AI-based HDBSCAN algorithm

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "title": "Stage 1 - Visual Temporal Reasoning",
  "description": "Reason step-by-step to identify urban transformations from time-sequenced
images.",
  "type": "object",
  "context": "Input is a collage of sub-images showing the same building across years, ordered from
top-left (oldest) to bottom-right (most recent), with visible timestamps. Focus strictly on the
central building; disregard fences, walls, bridges, and other elements. Camera angles and lighting
may vary. Interpret only if more than 50% of the building is visible.",
  "properties": {
    "image_descriptions": {
      "type": "array",
      "description": "Chronological visual descriptions of sub-images, formatted as '<YEAR>:
<description>'.",
      "items": {
        "type": "string"
      },
      "minItems": 1,
      "maxItems": 6
    },
    "description": {
      "type": "string",
      "description": "Detailed analysis of changes observed in the central building over time."
    },
    "transformation_type": {
      "type": "array",
      "description": "Detected transformation types with associated years.",
      "items": {
        "type": "object",
        "properties": {
          "type": {
            "type": "string",
            "enum": [
              "Facade",
              "Structure",
              "Function",
              "No Building",
              "Insufficient Data"
            ]
          },
          "year": {
            "type": "integer",
            "description": "Year when the transformation occurred."
          }
        },
        "required": [
          "type",
          "year"
        ]
      }
    }
  },
  "required": [
    "image_descriptions",
    "description",
    "transformation_type"
  ],
  "instructions": [
    "- Only describe sub-images with visible timestamps.",
    "- Ignore fences, walls, bridges, and side structures; focus only on the main building.",
    "- Small changes like shutter closures, graffiti appearance/removal, or natural aging are NOT
transformations.",
    "- Definitions:",
    "  - 'No Building': No visible structure across all images.",
    "  - 'Insufficient Data': Only one timestamped image available.",
    "  - 'Facade': Significant surface alterations (material, color, signage, windows); excludes
natural aging.",
    "  - 'Structure': Major form changes (demolition, reconstruction, added floors).",
    "  - 'Function': Usage changes inferred from signage or internal layout; shutter closures or
graffiti do not indicate function change.",
    "- Maintain chronological reasoning based on timestamps."
  ]
}
```

Figure A.4: JSON schema used to guide visual temporal reasoning and transformation classification. The prompt defines structured keys for image descriptions, temporal synthesis, and multi-label transformation output.

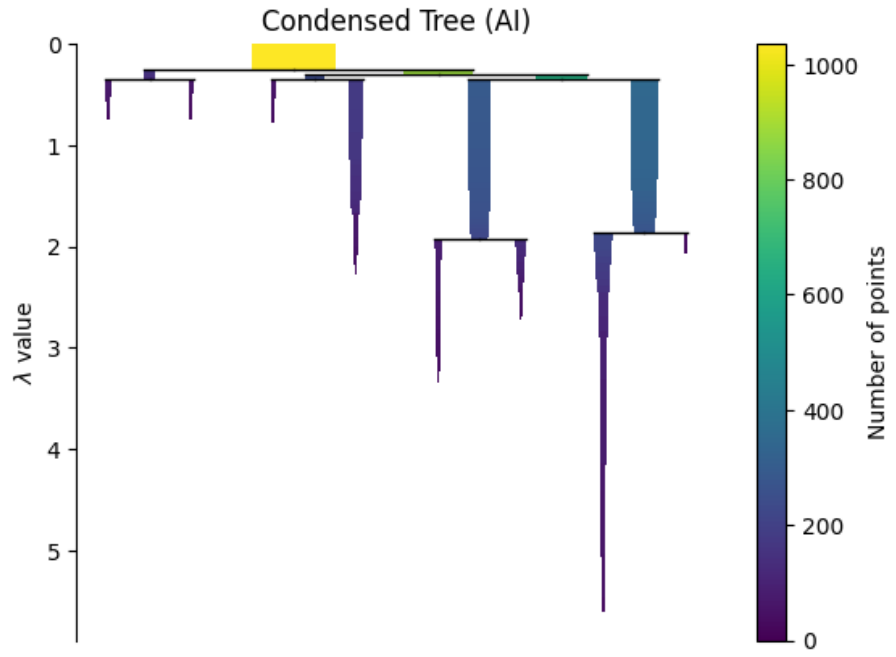


Figure A.5: Condensed tree visualization of the AI-based HDBSCAN clustering output

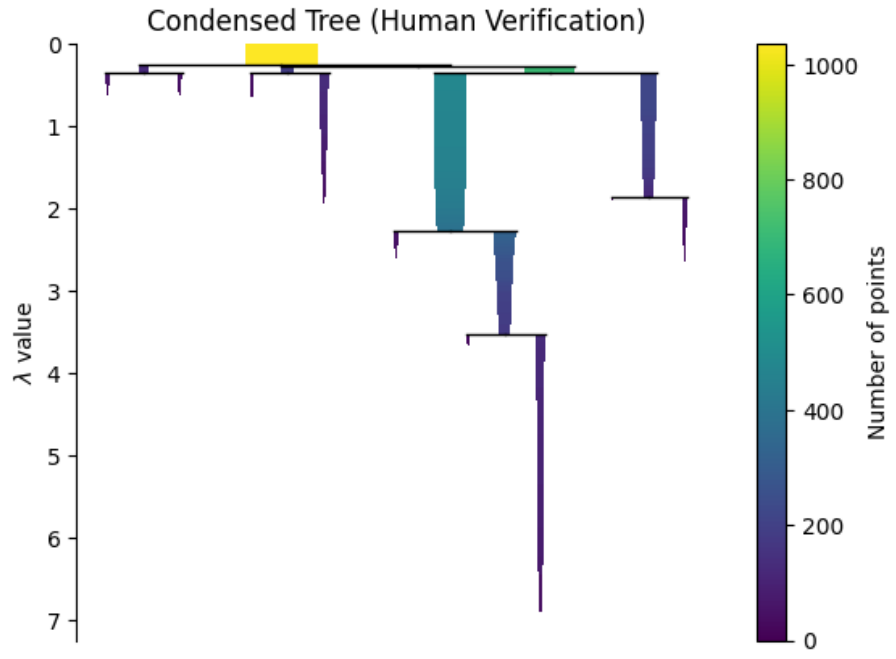


Figure A.6: Condensed tree visualization of the human-verified HDBSCAN clustering output.



Figure A.7: A comparison of GSV and Mapillary imagery demonstrates how user-generated content can reveal an intermediate facade state that is not captured by GSV. This highlights the potential of platforms like Mapillary in counter-cartographic practices, where unofficial or overlooked spatial conditions are made visible through alternative forms of documentation.