Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis
Data Science And Marketing Analytics

The View from the Street: How Urban Environmental Features Shape Housing Prices in Rotterdam

Linus Christian Wolff
(704716)

Supervisor:	Prof. Dr. Ir. Rommert Dekker
Second assessor:	Prof. Dr. Martijn G. de Jong
Date final version:	4th August 2024

Abstract

This study investigates the impact of urban environmental factors on residential property prices in Rotterdam, leveraging advanced computer vision techniques to extract features from street-level imagery. We extend traditional hedonic pricing models by incorporating visual characteristics of the urban environment. Analyzing 6,691 property listings and over 200,000 street view images, we employ semantic segmentation and object detection models to quantify urban greenery, pedestrian presence, and transportation infrastructure. Our methodology compares linear hedonic pricing models with non-linear Random Forest approaches across multiple spatial scales. Results demonstrate that incorporating image-derived environmental features significantly improves the predictive accuracy of property valuation models, with Random Forest models consistently outperforming traditional linear methods. The Urban Greenery Index exhibits the strongest positive influence on property values for houses but the weakest for apartments, while the presence of bicycles shows the largest overall positive effect. We observe non-linear relationships and threshold effects in the impact of environmental factors, highlighting the complexity of urban housing valuation. This research offers a novel framework for integrating visual environmental data into property valuation models, providing valuable insights for urban planning and real estate economics.

Keywords: computer vision, hedonic pricing models, random forest, real estate valuation, street-view imagery, urban environmental factors

Contents

1 Introduction

The housing market has long been a cornerstone of societal wealth and well-being. With many households holding the majority of their wealth in their primary residence, housing prices significantly influence the quality of life for a large portion of the population. This economic reality has sparked interest from various stakeholders, including homeowners, insurers, policymakers, and real estate tax assessors. Consequently, researchers have devoted considerable attention to developing accurate property valuation methods, with hedonic pricing models (HPM) emerging as a predominant approach (Rosen, 1974). These models estimate a property's value by analyzing the contributions of its individual attributes to the overall price, providing a detailed understanding of the factors driving market values.

In evaluating properties, the considerations of prospective homeowners extend far beyond the physical attributes of a house. They consider into a multitude of factors that impact the desirability and perceived value of a home. Guite et al. (2006) finds that the urban environment directly impacts social, economic, and health outcomes, making its quality a significant determinant in property values. Traditional HPMs have primarily focused on easily quantifiable structural attributes such as size, age, and the number of rooms (Chau and Chin, 2003). However, this approach overlooks critical environmental factors that significantly influence a property's desirability and, consequently, its market value (Chen et al., 2020). Elements such as amenities, neighborhood aesthetics, and the overall ambiance of a neighborhood contribute to how safe and desirable it feels (Law et al., 2019).

Some environmental factors, such as air and water quality, have been included in hedonic pricing models since the late 1960s and are consistently found to significantly impact property values (Boyle and Kiel, 2001). However, many critical environmental factors like traffic levels and noise pollution, are very challenging to quantify (Zhang and Dong, 2018). Recent technological advancements, particularly in machine learning and computer vision, have opened new avenues for incorporating these previously difficult-to-quantify environmental factors into real estate valuation models. The proliferation of unstructured data, such as images and text, has revolutionized our ability to capture and analyze complex environmental attributes (Potrawa and Tetereva, 2022). Various approaches have emerged to integrate image data into hedonic pricing models, ranging from classification and segmentation techniques to end-to-end methods that directly incorporate image data into pricing models (Poursaeed et al., 2018; Zhang and Dong, 2018).

This study builds upon these advancements, proposing a novel framework that leverages street-level imagery to extract and quantify urban environmental factors influencing housing prices in Rotterdam. Our approach offers an automated method to include proxies of human perception of urban surroundings, focusing on two key aspects: visual appeal (quantified through greenery) and neighborhood vitality (measured by the presence of people and vehicles). By quantifying these factors, we aim to capture elements of neighborhood desirability that may significantly influence property values.

Problem Statement

Despite the progress in real estate valuation methods, most studies still fail to adequately consider the broader environmental context. This oversight can lead to inaccurate valuations, as these models often do not capture the full spectrum of elements influencing buyer decisions. For instance, two properties with identical structural attributes may have significantly different market values based on their surrounding environment. A home situated near a serene park might command a premium compared to an otherwise similar property located beside a busy highway (Chen et al., 2020).

Recent studies have begun to address this gap by incorporating various types of unstructured data into valuation models. Images, in particular, have emerged as a valuable resource for capturing previously overlooked aspects of properties and their surroundings. Approaches range from analyzing interior images to assess the quality of materials and design (Poursaeed et al., 2018), examining exterior property images to assess the exterior appearance of a property (You et al., 2017), utilizing street view images for neighborhood characteristics (Zhang and Dong, 2018), and employing aerial imagery for broader urban context (Law et al., 2019).

While these approaches have advanced our understanding, there remains a critical need for research specifically focusing on the impact of neighborhood-level visual characteristics on property prices. This study aims to address this gap by leveraging street-level imagery to extract and quantify urban environmental factors influencing housing prices in Rotterdam.

This study aims to address the central question:

> **How do urban environmental factors extracted from street-level imagery influence residential housing prices in Rotterdam?**

This research distinguishes itself from previous studies in several key aspects:

- Focus on Neighborhood Characteristics: Unlike studies that assess the impact of interior features (Poursaeed et al., 2018) or individual property exteriors (You et al., 2017), we concentrate on the visual characteristics of the surrounding neighborhood.
- Automated Extraction of Environmental Factors: We systematically extract and quantify environmental factors from street-level imagery. This approach offers an objective and scalable method for assessing neighborhood characteristics (Zhang and Dong, 2018).
- Multi-scale Analysis: We explore how the influence of environmental factors varies with distance from the property. This addresses the hypothesis proposed by Law et al. (2019) that buyers might value a visually desirable neighborhood more than just a visually appealing street. By using different radii to define neighborhood contexts, we aim to determine the spatial extent of environmental influences on property values.
- Comprehensive Set of Environmental Factors: Our study examines a range of factors including greenery, the presence of people, and the prevalence of vehicles. This combination allows us to capture aspects of both visual appeal and neighborhood vitality.

- Comparison of Linear and Non-linear Models: We employ both traditional linear hedonic pricing models and more advanced non-linear techniques, specifically Random Forest models. This dual approach enables us to capture potential non-linear relationships and threshold effects in the impact of environmental factors on housing prices, while also providing a basis for comparison with established methods.

We utilize state-of-the-art computer vision models for semantic segmentation to quantify greenery (Cheng et al., 2022) for object detection to count people, cars, and bicycles (Redmon and Farhadi, 2018). This approach allows us to investigate how the inclusion of urban environmental factors impacts the accuracy of housing price predictions and examine the specific effects of various environmental factors on property values.

To address our research question, we employ a comprehensive analytical framework. This approach begins with acquiring diverse data sources, namely property listings and street view imagery. We then extract environmental features from the street view images using the aforementioned techniques, which are subsequently integrated with property data through spatial interpolation to create a unified dataset. We employ both traditional linear hedonic models and more advanced Random Forest techniques to analyze their impact on housing prices, and finally conduct an in-depth interpretation of our results.

By addressing these aspects, our study aims to provide a nuanced and comprehensive understanding of how urban environmental factors influence property values. This research has the potential to enhance the accuracy of real estate valuation models, inform urban planning decisions, and provide insights into the complex relationship between neighborhood characteristics and housing prices.

Relevance

This research occupies a unique intersection of urban planning and real estate valuation, with potential contributions to both practical applications and theoretical frameworks in these fields. The findings could benefit various stakeholders, including municipalities, developers, and real estate professionals, by providing a more objective approach to property valuation.

In the Netherlands, this study has particular relevance to the Wet waardering onroerende zaken (WOZ, or Real Estate Valuation Act) system, which forms the basis for property taxation. Currently, this process often relies heavily on manual assessment and expert judgment. Our approach offers a potential path towards more objective and consistent property valuations, which could enhance the fairness and accuracy of property tax assessments.

For urban planners and policymakers, our research provides insights into how specific environmental factors contribute to neighborhood desirability and property values. This knowledge could inform strategies for urban development and regeneration, helping to create more attractive and valuable urban spaces. For instance, understanding the economic impact of green spaces could guide decisions on infrastructure investments and zoning policies.

Academically, this study extends the discourse on real estate valuation and hedonic pricing models by proposing a novel framework for incorporating urban environmental factors derived from unstructured data sources. Using advanced computer vision

techniques, we contribute to computational urban studies and bridge the gap between theoretical data science and practical applications in urban planning and real estate economics.

Outline

The remainder of this thesis is structured as follows: The **Literature Review** provides a comprehensive examination of previous research on real estate valuation, focusing on the evolution of hedonic pricing models and the increasing use of unstructured data in property valuation. It also explores the growing body of work on quantifying urban environmental factors and their impact on property values. The **Hypotheses** section outlines our expectations regarding the influence of environmental factors on housing prices, based on insights from the literature and the specific context of Rotterdam. The **Data** section details our datasets, including property listings from Funda.nl and an extensive collection of street view images. It explains how these diverse data sources are integrated to create a comprehensive picture of Rotterdam's housing market and urban environment. The **Methodology** section elucidates our analytical framework, describing both the linear hedonic and non-linear Random Forest models employed. It also provides a detailed explanation of the computer vision techniques used to extract environmental features from street-level imagery. The **Results** section presents our findings, comparing the performance of different models and analyzing the impact of various environmental factors on housing prices across different spatial scales. The **Conclusion** summarizes the key insights of our study, while the **Discussion** reflects on the implications of our findings for stakeholders in real estate and urban planning. It also considers the limitations of our approach and suggests directions for future research.

2 Literature Review

This literature review provides a comprehensive examination of the evolution of real estate valuation methods, with a focus on integrating urban environmental factors and applying advanced computational techniques. The review is structured to address key aspects related to our research question. We begin by exploring the foundations and evolution of the Hedonic Pricing Model, which forms the theoretical basis for our approach. We then examine the transition from linear models to machine learning approaches, highlighting the need for more flexible modeling techniques to capture complex relationships in housing markets. The review continues with an analysis of how unstructured data, particularly images, have been integrated into real estate valuation, directly informing our methodology for extracting environmental factors from street-level imagery. Finally, we discuss recent advancements in urban environmental factor analysis and computer vision techniques. Throughout the review, we trace the progression from traditional valuation methods to modern computational approaches, setting the stage for our study on housing prices in Rotterdam.

2.1 Hedonic Pricing Model: Foundations and Evolution

The Hedonic Pricing Model (HPM) serves as the basis of our approach to real estate valuation. The Hedonic Pricing Model (HPM) serves as the cornerstone of our approach to real estate valuation. Introduced by Lancaster (1966) and further developed by Rosen (1974), the HPM posits that the value of a good is derived from its constituent characteristics. This concept is particularly relevant to our study as it allows us to decompose property values into the contributions of various attributes, including the urban environmental factors we aim to quantify.

The Hedonic price modeling framework has been extensively applied to real estate valuation, where property characteristics are categorized into locational, structural and neighborhood attributes (Chau and Chin, 2003). This categorization informs our approach, especially in how we differentiate between property-specific features and broader environmental factors.

Structural and Locational Attributes

While our study focuses on neighborhood-level environmental factors, understanding the role of structural and locational attributes is crucial for developing a comprehensive valuation model. Studies by Rodriguez and Sirmans (1994), Fletcher et al. (2000), and Garrod and Willis (1992) have consistently shown positive correlations between property values and features such as floor area, number of rooms, and overall quality. For instance, Garrod and Willis (1992) found that an additional room can increase a property's value by about 7%. Additional features such as building age, which negatively impacts

property values (Clark and Herrin, 2000), and the existence of garages and heating systems also have a significant influence (Forrest et al., 1996; Garrod and Willis, 1992; Michaels and Smith, 1990). These findings underscore the importance of controlling for these variables in our model to isolate the effects of environmental factors.

Locational factors have also received considerable research attention. Accessibility, particularly to the Central Business District (CBD), has been a cornerstone of many studies (McMillan et al., 1992; Palmquist, 1989). Research consistently shows that improved accessibility generally correlates positively with property values (So et al., 1997). The work of Mok et al. (1995) and Rodriguez and Sirmans (1994) on the impact of views on property values is particularly relevant to our research. Their findings that attractive views can significantly increase property values (by up to 8% in some cases) support our hypothesis that visual environmental factors play a crucial role in determining housing prices.

Beyond geographical attributes within a city, locational attributes also encompass broader regional or cultural factors. For instance, Bourassa et al. (1999) found that houses with 'lucky numbers' had a significant positive impact on sale prices. Chau et al. (2001) observed a comparable effect in Hong Kong, although the numbers deemed fortunate differed due to cultural variations.

Neighborhood Attributes and Environmental Factors

The importance of neighborhood attributes in property valuation, as highlighted by Goodman (1978) and Linneman (1980), validates our focus on environmental factors at the neighborhood level. Their work demonstrates that a substantial portion of variation in property values can be attributed to neighborhood characteristics, especially for structurally similar houses.

Researchers have typically categorized neighborhood attributes into several key areas: socio-economic, local government services and externalities (Chau and Chin, 2003). Socio-economic variables, such as the income of the residents within a neighborhood, have been shown to influence property values (Richardson et al., 1974). Local government services, particularly the quality of public schools, have also been found to possess a significant impact on house prices (Clark and Herrin, 2000).

Externalities also play a significant role in shaping neighborhood desirability. Crime rates have been consistently shown to negatively affect property values, with Clark and Herrin (2000) finding that property prices in California were significantly negatively correlated with homicide rates. Similarly, various forms of noise and air pollution, have been extensively studied in the context of property valuation. Research on traffic noise (Palmquist, 1989) and airport noise (Espey and Lopez, 2000) has consistently demonstrated how environmental disturbances can negatively impact the market value of residential properties. A similar adverse effect on property values has been observed in studies examining air pollution levels (Harrison Jr and Rubinfeld, 1978).

Environmental quality, including pleasant surroundings, quietness, unpolluted air and water, among others, has gained increasing attention in recent years as a crucial neighborhood attribute. Tyrväinen (1997) found that green housing districts and accessibility to forested recreation areas command higher prices among urban residents in Finland. Their work offers a historical perspective on quantifying environmental factors. Their

use of manual methods to measure green areas contrasts sharply with our computer vision approach, highlighting the technological advancements that enable our study, which will be detailed later.

2.2 From Linear Models to Machine Learning Approaches

The majority of the previously mentioned studies employed the standard linear hedonic pricing model, which utilizes ordinary least squares (OLS) regression. The simplicity in formulating, estimating, and interpreting these models allows for straightforward interpretation of the attributes comprising a good.

However, the conventional hedonic pricing model has faced criticism for imposing strong assumptions that may not always align with real-world complexities. While traditional linear hedonic models have been widely used, their limitations in capturing complex, non-linear relationships in real estate markets have been increasingly recognized. Chau and Chin (2003) and Malpezzi et al. (2003) critique the assumptions of perfect competition and market equilibrium inherent in linear models, highlighting the need for more flexible approaches.

The issues of multicollinearity and interaction effects in linear models further justify our use of advanced modeling techniques. As Butler (1982) pointed out, many housing characteristics are inherently related. For instance, the number of bedrooms in a house is typically not independent of the number of bathrooms. So et al. (1997) demonstrated this in Hong Kong's housing market, showing that an apartment's floor level was significant mainly due to its interaction with the quality of the view. Our approach, particularly the use of Random Forest models, aims to capture these complex interactions more effectively.

Studies by Potrawa and Tetereva (2022) and Chen et al. (2020) revealing non-linear relationships between living space, greenery, and property values are further examples that housing characteristics may have threshold effects or diminishing returns. Potrawa and Tetereva (2022) found diminishing returns in the relationship between living area and property value, with marginal value dropping significantly after 140 square meters – a nuance missed by linear models. Similarly, Chen et al. (2020) discovered a non-linear relationship between neighborhood greenery and property values, with positive effects only emerging after a certain threshold.

These complexities have led researchers to explore non-parametric approaches. Mason and Quigley (1996) were among the first to employ a generative additive model, successfully identifying non-linearities in housing data. Subsequently, numerous studies have utilized various machine learning algorithms.

Comparative studies have demonstrated the superior predictive performance of machine learning models. Hong et al. (2020) compared random forest (RF) models to traditional OLS regression using apartment transaction data in South Korea, finding the RF model significantly outperformed OLS, with an average percentage deviation of approximately 6% compared to 20% for OLS. Kostic and Jevremovic (2020), Mu et al. (2014), and Abidoye and Chan (2018) showed similar improvements using gradient boosting models, support vector machines, and artificial neural networks, respectively, providing a strong rationale for our use of Random Forest models alongside traditional hedonic pricing models.

However, these studies primarily focused on predictive performance, often at the expense of interpretability, which is a key objective in hedonic modeling. To address this, we incorporate recent developments in model interpretation techniques, such as SHAP (Shapley Additive exPlanations) values. Doan et al. (2024) employed this technique to analyze the non-linear effects of air pollution on housing prices in New York City, revealing that pollution levels had a negligible impact on prices when they were below the city average. Similarly, Chen et al. (2020) applied SHAP analysis to quantify the threshold at which greenery began to positively impact property values.

2.3 Integrating Unstructured Data in Real Estate Valuation

The real estate sector has begun to harness the potential of unstructured data, especially images, to enhance valuation models. The growing interest in leveraging unstructured data for real estate valuation is driven by the recognition that traditional structured data often fails to capture the nuanced environmental and aesthetic factors that significantly impact property desirability and, consequently, its market value (Law et al., 2019).

Recent studies have demonstrated various approaches to incorporating image data into valuation models, each focusing on different aspects of visual information. Potrawa and Tetereva (2022) used images looking out from properties, finding that attractive views (e.g., city or park) significantly predicted rent prices. Poursaeed et al. (2018) improved predictions by classifying interior images by luxury level. While valuable, these approaches do not address the broader environmental context that our study aims to capture.

More aligned with our research is Zhang and Dong (2018)'s use of street-view images to measure the impact of visible greenery on housing prices in Beijing. Their development of a specific greenery index integrated into a hedonic regression model serves as a key precedent for our study. We expand upon this approach by considering a wider range of environmental factors, aiming to provide a more holistic understanding of how urban environmental characteristics influence property values.

Yencha (2019) further illustrated the versatility of street-level imagery by assessing walkability and its impact on home prices. Their use of computer vision to identify pedestrian infrastructure aligns with our methodological approach. Similarly, Law et al. (2019)'s pipeline for automatically extracting visual features from street-level and aerial imagery provides a methodological foundation, though our approach differs in focusing on specific, interpretable environmental factors rather than using an end-to-end model.

Beyond real estate valuation, studies by Garrido-Valenzuela et al. (2023) and Gebru et al. (2017) demonstrate the broader potential of street-view imagery in urban analysis. Garrido-Valenzuela et al. (2023) used such images to monitor urban space usage and assess urban interventions in the Netherlands, finding correlations between urban density and features like block size and the presence of food establishments and bicycles. Gebru et al. (2017) employed deep learning on street-view images to estimate neighborhood demographics across American cities, revealing correlations between vehicle types and socio-economic indicators.

While these latter studies do not directly address real estate valuation, they highlight the wealth of socio-economic and urban planning insights that can be extracted from street-level imagery. This showcases the potential of such data in broader urban analysis

contexts and underscores the value of our approach in leveraging this rich data source for property valuation.

The Digital Transformation of Real Estate Marketing

While our study primarily focuses on the valuation aspects of real estate, we recognize the evolving landscape of real estate marketing and its implications for our research. The real estate industry has undergone significant transformations in recent decades, particularly in how properties are marketed and how information is conveyed to potential buyers.

Hamilton and Gunesh (2003) highlights a pivotal shift in the real estate industry from traditional people-mediated interfaces towards a business model that relies heavily on technology-mediated interaction with customers. As Shaw (2020) argues, these digital platforms are fundamentally changing how market participants connect and interact. The adoption of digital technologies in real estate marketing has been rapid and widespread. Bond et al. (2000) found that as early as 2000, the majority of real estate firms operated their own websites. Since then, driven primarily by the advent of digital technologies, the evolution of real estate marketing over the past twenty years has fundamentally transformed the industry (Petermann, 2021), with a shift towards the dominance of larger aggregated property listing sites.

The shift towards digital platforms in real estate marketing has significant implications for our research on urban environmental factors. Shaw (2020) suggests that future real estate markets will increasingly be defined by how online platforms connect market participants through a digital stack of technology. Moreover, Sihi (2018) utilizes VR and AR experiences to highlight how digital technologies can enhance the purchase experience by providing consumers more knowledge and understanding about a product or service before actually seeing or experiencing it. Petermann (2021) notes that the future of real estate marketing lies in the integration of even more technologies into the online marketing presence. Our extraction of objective, data-driven insights about neighborhood characteristics from street-level imagery exemplifies this trend. It could enhance the information content available on these platforms, potentially providing a virtual assessment of the neighborhood before physically visiting a property.

2.4 Quantifying Urban Environmental Factors through Computer Vision

Urban environmental factors encompass a wide range of characteristics pertaining to the surrounding area of a property, providing insights into neighborhood quality, aesthetics, and overall livability. Unlike studies that focus on interior images of properties (e.g., Poursaeed et al., 2018) or those that quantify views from specific apartments (e.g., Jim and Chen, 2006), urban environmental factors are characteristics of the neighborhood or surrounding area, rather than attributes of the property itself or large-scale regional features.

A growing body of literature leverages various types of imagery to quantify urban environmental factors. Studies by Zhang and Dong (2018), Fu et al. (2019), Chen et al. (2020), and Ye et al. (2019) demonstrate the effectiveness of semantic segmentation techniques in measuring factors like greenery, sky view, and building density. These

studies inform our choice of environmental factors and support our hypothesis that they significantly influence property values.

Zhang and Dong (2018) and Fu et al. (2019) used street-view images to measure visible greenery and sky view, finding significant positive effects on real estate values. Chen et al. (2020) expanded this approach by integrating both street-level and satellite imagery, while Ye et al. (2019) reinforced the positive correlation between street greenery and property values across different urban contexts. Traditional approaches using Geographic Information Systems (GIS) data have also been employed. Kolbe and Wüstemann (2014) uses GIS data to examine the impact of water coverage and green spaces on apartment prices in Cologne, Germany, providing a valuable comparison point for our image-based approach. Their findings align with the image-based studies, showing positive effects of these features on property values.

Our research extends beyond these traditional factors to include measures of urban vitality. While studies such as Zhao et al. (2023) have relied on human volunteers to assess perceptions of urban environments (e.g., liveliness, friendliness, interest), we aim to capture similar information through automated, objective means. By quantifying the presence of people, cars, and bicycles in street scenes, we gauge neighborhood vitality, traffic levels, and mobility options. This novel approach advances the field by offering an objective and scalable method for assessing these urban characteristics.

Some researchers have taken a more holistic approach using satellite imagery. Semnani and Rezaei (2021) demonstrated improved prediction accuracy by incorporating satellite images into housing price models. Others, like Muhr et al. (2017) and Bency et al. (2017), used satellite imagery to extract specific environmental features and explore their spatial scale. For instance, Bency et al. (2017) employed satellite images at multiple zoom levels, suggesting that larger neighborhood sizes more accurately predict property values. These studies motivate our decision to analyze environmental factors at various radii from each property.

The evolution of these approaches is closely tied to advancements in computer vision and machine learning techniques, which have dramatically expanded our ability to quantify and interpret visual data from urban environments. The foundations of modern computer vision are rooted in deep learning, a subset of machine learning that uses artificial neural networks inspired by the human brain to learn from large amounts of data. From the introduction of one of the first Convolutional Neural Networks (CNNs) by LeCun et al. (1998), demonstrating the successful application of CNNs to handwritten character and word recognition, to the development of more complex architectures like AlexNet (Krizhevsky et al., 2012), VGG (Simonyan and Zisserman, 2014), and GoogLeNet (Szegedy et al., 2014), algorithmic and computational advancements have set the stage for more specialized applications. The introduction of the Transformer architecture (Vaswani et al., 2017) and its adaptation to vision tasks (Dosovitskiy et al., 2020) further expanded the field's capabilities.

For our specific needs, segmentation and object detection techniques are particularly relevant. The introduction of SegNet by Badrinarayanan et al. (2016) represented a major advancement in semantic segmentation, enabling pixel-wise classification of images. This capability was quickly applied in real estate valuation research, as demonstrated by Zhang and Dong (2018) in their analysis of street-visible greenery's impact on housing prices. Our study leverages two state-of-the-art models: YOLO (You Only Look Once) for object detection (Redmon et al., 2016) and Mask2Former for semantic segmentation (Cheng et al., 2022). YOLO's efficiency in processing large volumes of images, even on

consumer hardware, and Mask2Former's accuracy in delineating complex urban features make them ideal for our analysis.

The effectiveness of these advanced models is enhanced by comprehensive datasets like Cityscapes (Cordts et al., 2016), a large-scale benchmark dataset for urban scene understanding, complete with pixel-level annotations for thousands of images collected from a variety of cities. By using models pre-trained on Cityscapes, we ensure that our feature extraction techniques are robust and applicable to the diverse urban landscape of Rotterdam.

Our approach of extracting specific, interpretable environmental factors distinguishes our study from end-to-end methods that derive latent features directly from images (Semnani and Rezaei, 2021; Yazdani and Raissi, 2023). This method allows us to maintain interpretability in our results, crucial for translating our findings into insights for urban planning and real estate valuation.

This review reveals several key insights that inform our research. It underscores the importance of neighborhood-level environmental factors in determining property values and highlights the potential of unstructured data in enhancing their valuation. It also showcases the limitations of traditional linear hedonic models in capturing complex relationships in real estate markets, supporting our use of more advanced machine learning techniques. However, it also reveals a gap in the literature: while several studies have incorporated image data into valuation models, few have focused on extracting and quantifying multiple urban environmental factors from street-level imagery at various spatial scales. Our study addresses this gap by leveraging computer vision to extract a comprehensive set of environmental factors from street-view images of Rotterdam. By doing so, we contribute to the evolving field of computational urban studies and provide a framework for incorporating visual environmental data into real estate valuation models.

3 Hypotheses

Based on existing literature, empirical observations, and logical reasoning concerning urban environments and real estate valuation, we propose the following hypotheses to guide our investigation into how urban environmental factors influence housing prices in Rotterdam:

3.1 Enhanced Predictive Power of Environmental Factors

The inclusion of urban environmental factors extracted from street-level imagery will lead to more accurate predictions of housing prices compared to models using only traditional structural and locational attributes.

While these factors are crucial, they do not capture the full range of elements that affect property values. Environmental factors such as the presence of greenery, traffic levels, and indicators of urban vitality significantly influence buyer preferences and, consequently, housing prices.

This hypothesis is supported by several studies in the literature. For instance, Chen et al. (2020) demonstrated that incorporating environmental features derived from street-view and satellite imagery significantly improved the predictive accuracy of real estate models in Shanghai. Similarly, Law et al. (2019) found that augmenting traditional housing attribute models with visual data enhanced house price estimations in London.

3.2 Directional Influence of Environmental Features

Specific environmental factors extracted from street-level imagery will have directional effects on housing prices, with these effects potentially differing between houses and apartments.

1. **Greenery Index:** Higher levels of visible greenery are expected to positively affect property values, with a more pronounced effect for houses compared to apartments. Studies by Zhang and Dong (2018) and Jim and Chen (2006) have consistently found that properties with higher levels of visible greenery command higher prices. The stronger effect on houses is expected because they are typically located in less dense areas, where inhabitants place greater value on the surrounding nature.
2. **Presence of People:** A higher count of people is hypothesized to have a positive effect on property values, because the presence of people can indicate the vitality and vibrancy of an area. This aligns with Garrido-Valenzuela et al. (2023), who found that urban density correlates with various positive urban characteristics. We expect this effect to be more pronounced for apartments, which are typically located in more densely populated, urban areas where vibrancy might be more highly valued.

3. **Number of Cars:** A higher number of cars is expected to negatively impact housing prices, with potentially stronger effects for houses. High traffic levels are associated with noise, air pollution, and reduced safety, making neighborhoods less desirable. This is supported by studies such as Espey and Lopez (2000), which found noise significantly negatively impacts prices in less dense areas. Conversely, Jim and Chen (2006) found noise had no impact on property values in a densely populated area in China. However, we acknowledge the potential for a non-linear relationship, as a moderate number of cars might indicate good accessibility or commercial activity, which could be valued positively, especially for apartments.
4. **Number of Bicycles:** A higher count of bicycles is hypothesized to have a positive impact on property values, particularly for apartments. Garrido-Valenzuela et al. (2023) found that the presence of bicycles is an indicator of a vibrant area. Additionally, it may serve as a proxy for good pedestrian and cycling infrastructure, which could be positively valued. We expect this effect to be more pronounced for apartments, which are often located in more central, bicycle-reliant urban areas.

These directional effects might vary depending on local cultural and socio-economic contexts. For instance, in some densely populated urban areas of Rotterdam, high traffic might be less of a deterrent if the neighborhood is otherwise highly desirable.

3.3 Spatial Decay of Environmental Influences

The impact of environmental factors on housing prices is expected to decrease with increasing distance from the property.

We anticipate that the influence of environmental features will be more pronounced when they are in close proximity to a property. This hypothesis aligns with findings from Ye et al. (2019), who observed that the impact of green spaces on property values is more significant within closer proximity.

To test this hypothesis, we analyze environmental factors at multiple radii (250m, 500m, and 1000m) around each property. This multi-scale approach allows us to capture the spatial decay of environmental effects and identify the most relevant neighborhood size for different factors. Our expectation of decreasing influence with distance is based on the assumption that potential buyers are more concerned with the immediate surroundings of a property.

However, we acknowledge that this relationship may not be universal and can depend on the environmental features studies. As noted by Bency et al. (2017), who explored the spatial scale of environmental factors using satellite images at multiple zoom levels, larger neighborhood contexts can provide better predictions of property values. Certain features, such as large parks or significant water bodies, may have a broader radius of influence, affecting property values over larger distances.

3.4 Identifying Non-Linear and Threshold Effects in Property Valuation

The relationships between environmental factors and housing prices are expected to exhibit non-linear patterns and threshold effects.

Real estate markets are complex systems, and the impact of environmental factors on housing prices is unlikely to be straightforward or linear. We anticipate non-linear relationships and potential threshold effects for several reasons:

1. **Greenery:** Based on findings from Chen et al. (2020), we expect that the positive effect of greenery on property values may exhibit diminishing returns beyond a certain level.
2. **Traffic (Cars):** We hypothesize that the impact of traffic on property values may follow a threshold pattern. A low to moderate number of cars might not significantly affect prices, but beyond a certain threshold, increased traffic could lead to a sharp decline in property values due to noise, pollution, and safety concerns. This effect may be more acute for houses, which are often valued for their quieter surroundings.
3. **Urban Vitality (People and Bicycles):** The relationship between these indicators of urban vitality and property values may follow an inverse U-shaped curve. Very low levels might indicate a lack of amenities or accessibility, moderate levels could be optimal, while very high levels might suggest overcrowding or over-commercialization. Different patterns might emerge for apartments and houses.

These types of effects are supported by various studies in the literature. For instance, Doan et al. (2024) found non-linear effects of air pollution on housing prices in New York City, revealing that pollution levels had a negligible impact on prices when they were below the city average, but their influence increased gradually when exceeding this threshold. Potrawa and Tetereva (2022) observed diminishing returns in the relationship between living area and property value beyond a certain size.

The hypotheses guide our investigation into the direct effects of environmental factors, their spatial influence at various scales, and potential non-linear relationships. By examining these aspects for both apartments and houses, we aim to uncover how the valuation of environmental characteristics might differ based on property type. This approach acknowledges that apartment dwellers and house owners may prioritize and value urban environmental features differently.

4 Data

4.1 Study Area

This study focuses on Rotterdam, the second-largest city in the Netherlands, as its area of investigation.

Rotterdam's urban landscape is characterized by a distinctive blend of modern architecture and historical elements, providing a diverse range of urban environments within a single metropolitan area. This diversity is crucial for our study, as it allows us to examine how different urban contexts within the same city affect property values. The city's housing market offers a broad spectrum of property types, from high-rise apartments in the city center to single-family homes in residential neighborhoods, enabling a comprehensive analysis of environmental factors across various property types and locations.

The city's geography features significant environmental diversity, including substantial green spaces, numerous water bodies (most notably the Nieuwe Maas river), and varying urban densities. This variability in urban form and environmental features makes Rotterdam an excellent case study for investigating the impact of urban environmental factors on housing prices.

Data availability was a crucial consideration in selecting Rotterdam for this study. The city benefits from comprehensive and accessible datasets, including a large number of property listings from Funda.nl, the Netherlands' largest real estate website. Additionally, through collaboration with local researchers, this study has access to an extensive street view image dataset of Rotterdam. This allows for a detailed exploration of the city's visual landscape as experienced by its residents and visitors.

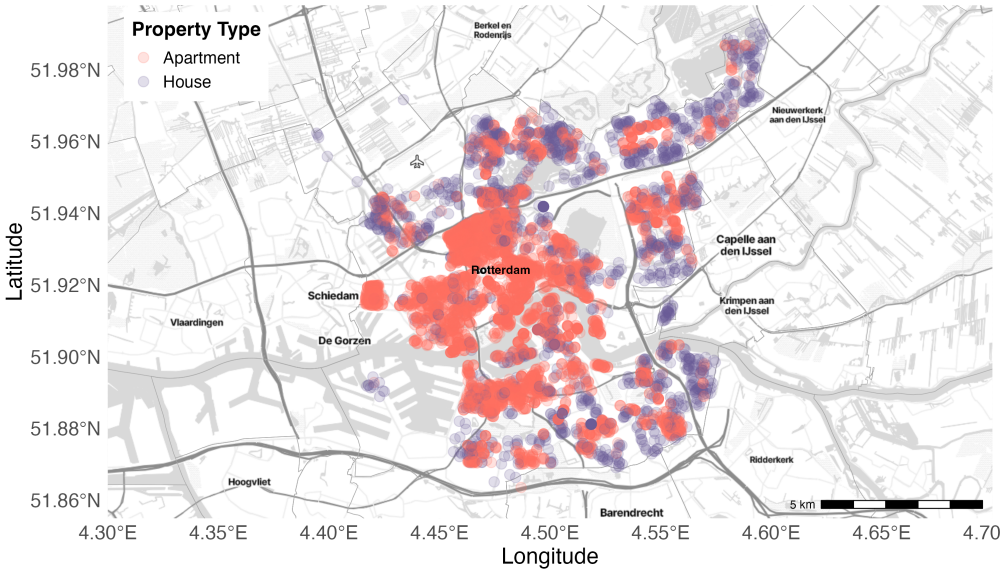

Figure 4.1: Property Listings Colored by the Property Type.

As illustrated in Figure 4.1 , the property listings in our dataset are distributed across central Rotterdam, ensuring comprehensive coverage of the city's diverse neighborhoods while maintaining a focus on the more densely populated urban core.

The choice of a Dutch city makes our findings particularly relevant in the context of governmental property valuations. As previously mentioned, municipalities annually determine the Wet waardering onroerende zaken (WOZ) value, which serves as the basis for various taxes. While we've previously discussed the WOZ's role in property taxation, it's worth emphasizing how our research methodology aligns with potential improvements to this system. By quantifying the impact of environmental features on property values using an automated approach, our study offers a practical demonstration of how municipalities might enhance their valuation processes. This is especially pertinent for Rotterdam, a city known for its innovative urban planning and development strategies. Our findings could provide local policymakers with empirical evidence to support more nuanced, objective assessments of property values, potentially leading to more equitable taxation and informed urban planning decisions.

4.2 Property Listings

The foundation of this study is a comprehensive dataset of property listings scraped from Funda.nl, the Netherlands' preeminent real estate platform. Funda's market dominance is significant: it attracts 4-6 times more unique visitors than its closest rivals and is majority-owned by NVM (Dutch Association of Real Estate Agents), which represents approximately 60% of estate agents in the Netherlands. This ownership structure gives Funda a substantial advantage in terms of listings. According to Niels (2018), 87% of house sellers consider advertising on Funda crucial, and an equal percentage of home-buyers use it as their primary search platform. This market position ensures our dataset offers a highly representative view of Rotterdam's housing market.

Our dataset comprises 6.691 unique property listings in Rotterdam, collected between March and May 2024, representing past listings of properties where the sale has already

occurred. This includes 5.009 apartments and 1.682 houses, with a mean property price of €417749 (apartments averaging €367.786 and houses €566.539). Each listing contains a rich set of variables primarily representing structural characteristics, aligning with one of the three main categories of housing attributes defined by Chau and Chin (2003). These variables include, but are not limited to, number of bedrooms, bathrooms, year of construction, energy label, and precise geolocation. Table 4.1 provides a detailed breakdown of the dataset:

Table 4.1: Overview of Structural Housing Attributes Differentiated by House Type.

Variable	**Apartments**, N = 5009 (75%)	**Houses**, N = 1682 (25%)
Price (€)	367786 (335000); [75000, 1895000]	566539 (485000); [150000, 2495000]
Living Area (m²)	85 (80); [15, 342]	135 (127); [54, 461]
Rooms	3 (3); [1, 9]	5 (5); [2, 14]
Construction Year	1961 (1957); [1855, 2026]	1969 (1978); [1649, 2025]
Energy Label		
Higher (A+, A, B, C)	3176 (63%)	1290 (77%)
Lower (D, E, F, G)	1833 (37%)	392 (23%)
Insulation		
Double Glass	3031 (61%)	897 (53%)
Fully Insulated	1012 (20%)	454 (27%)
Other	966 (19%)	331 (20%)
Heating		
Central Heating	2925 (58%)	1208 (72%)
District Heating	1106 (22%)	321 (19%)
Other	978 (20%)	153 (9%)

[1] Mean (Median); [Range]; n (%)

To ensure data integrity, we applied several processing steps. Only listings with complete information for all variables were included. Properties with asking prices below €50,000 were excluded as they typically represented non-residential structures, and an upper price cap of €2.500.000 was applied to eliminate rare outliers. All properties were successfully geocoded using the Google Maps API, providing precise latitude and longitude coordinates for each listing.

While our dataset provides rich insights into Rotterdam's real estate market, it's important to acknowledge its limitations. The property prices in our dataset represent the last known listing prices, which may not always reflect the final transaction values. This potential discrepancy between listed and actual sale prices introduces a degree of uncertainty into our analysis. Additionally, as web-scraped data, the listings may occasionally contain inaccuracies inherent in online platforms.

It's worth noting that access to official transaction data from the Kadaster (the Dutch land registry) would provide a more precise picture of actual sale prices. While such data was not available for this study, future research could benefit from incorporating Kadaster records to further validate and refine the findings presented here. In the literature, both types of data - official transaction records and listing prices - are commonly used. For instance, Hong et al. (2020) utilizes governmental transaction records for apartments in South Korea, covering about 40% of all apartment transactions. On the other hand, numerous studies have successfully employed real estate listing data, similar to ours (Potrawa and Tetereva, 2022; Ye et al., 2019; Zhang and Dong, 2018).

Despite these constraints, our dataset aligns well with the hedonic modeling context of this study, capturing the market's perception of property values based on the attributes visible to consumers during their home search process.

4.3 Street View Images

To enrich our analysis of Rotterdam's real estate market and address our research question on urban environmental factors, we incorporate a vast dataset of street view images (SVI). These images provide a visual representation of the urban environment, allowing us to quantify various aspects of the streetscape that may influence property values. As noted by Ye et al. (2019), SVI have been widely accepted as an effective means to quantify the built environment of streets, offering insights that traditional data sources often miss.

Our SVI dataset, initially sourced from Google Street View (GSV), was obtained through collaboration with researchers who had previously utilized this data (Garrido-Valenzuela et al., 2023). The image collection process followed a systematic approach to ensure comprehensive coverage of Rotterdam. A grid of points, each separated by 50 meters, was overlaid on the city map. This grid spacing is consistent with methods used in similar studies, such as Chen et al. (2020), ensuring a balance between granularity and computational feasibility.

For each grid point, a unique 360-degree panorama was captured at street level, which was then divided into four individual images: front, back, and two side views. This division allows for a more regular angle of view in the images subject to analysis. The resulting dataset consists of RGB images, each with a resolution of 900x600 pixels.

To align with our property listings data and minimize seasonal biases, particularly for metrics like the greenery index, we applied several filters to the initial dataset:

1. We restricted our analysis to images taken between June and September, ensuring consistent representation of vegetation.
2. Only images from 2017 or newer were included, better aligning with the timeframe of our property listings.
3. We focused on images captured during daylight hours to ensure visibility and consistency.

These filtering steps, while necessary for data quality, do introduce certain limitations. For instance, the Google Maps API does not provide granular timing information, preventing us from differentiating between weekdays and weekends. This could potentially bias calculations of metrics such as the number of cars present in images.

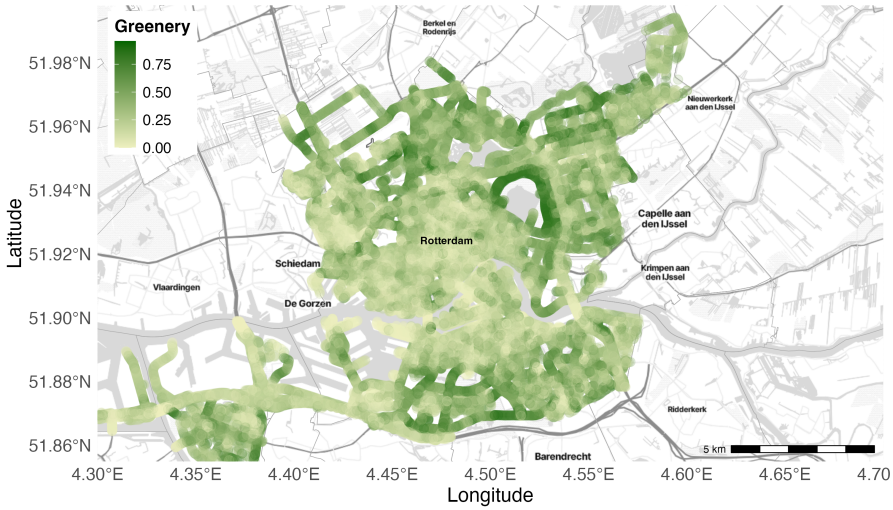

Figure 4.2: Distribution of Greenery across Rotterdam.

After applying these filters, our final dataset comprises 57,495 unique panoramas, corresponding to 229,980 individual images. Figure 4.2 provides a map illustrating the distribution of these images across Rotterdam, demonstrating comprehensive coverage of the city, particularly in areas where our property listings are located.

To further ensure data quality, we implemented a two-stage cleaning process. First, we employed CNN classifiers to automatically identify and remove invalid images, such as those that were too dark. This was followed by a manual review process to catch any remaining issues, such as images where large vehicles blocked the entire view. This approach is similar to that used by Law et al. (2019) in their study of London's housing market, ensuring the reliability of our visual data for subsequent analysis. Figure 4.3 presents examples of both valid and invalid images from our dataset.

Figure 4.3: Example Street View Images: Valid (top) and Invalid (bottom).

While this extensive dataset provides a rich source of information, it also presented significant computational challenges. The initial collection of over one million images occupied more than 100GB of storage, making it infeasible to run complex models on the entire set. Our filtering and cleaning processes not only improved data quality but also reduced computational demands to a manageable level.

In the subsequent analysis, these images will be related to our property listings through a process of spatial interpolation, which will be detailed in the methodology section.

5 Methodology

5.1 Research Design and Analytical Approach

Our research design integrates traditional real estate valuation techniques with advanced computer vision and machine learning methods, allowing for a comprehensive analysis of both structural and environmental determinants of property values. The analytical framework of this study can be conceptualized in five main stages. This multi-stage process, illustrated in Figure 5.1, enables us to systematically address our research questions and test our hypotheses regarding the impact of urban environmental factors on housing prices.

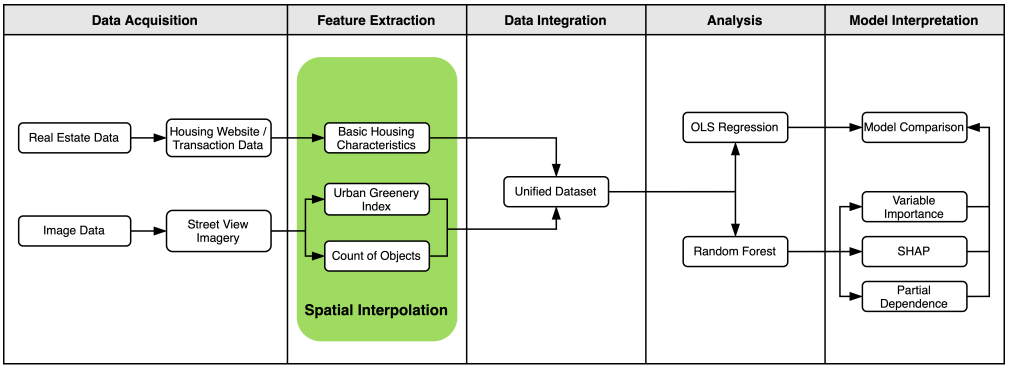

Figure 5.1: Research Design Framework.

The analytical framework consists of five stages:

1. **Data Acquisition:** Collecting property listings and street view images.
2. **Feature Extraction:** Deriving environmental variables from street view images using computer vision techniques.
3. **Data Integration:** Combining extracted features with property data into a unified dataset through spatial interpolation.
4. **Modeling:** Employing both linear hedonic and Random Forest models to analyze the impact of environmental factors on housing prices.
5. **Interpretation:** Assessing model performance and interpreting results using various techniques.

The following sections will detail the components of our methodology. We begin with our approach to visual feature extraction, a critical and innovative aspect of this study. Following the visual feature extraction, we will discuss our modeling approaches, comparing traditional linear hedonic models with more advanced machine learning techniques. We then outline our methods for model assessment and interpretation, which enable us to

evaluate the performance of our models and gain insights into the relative importance of different factors in determining property values.

5.2 Extracting Visual Environmental Features

Our study analyzes Rotterdam's housing market dynamics by extracting four key urban environmental features from street-level imagery: greenery, people, cars, and bicycles. These features, selected based on existing literature and our research objectives, aim to capture both the physical and social aspects of urban environments that influence property values. The pre-processed street-view images, as detailed in the Data section, form the basis of our analysis.

Quantifying Urban Greenery through Semantic Segmentation

The quantification of urban greenery from street-level imagery forms a crucial component of our study, requiring us to accurately segment and classify vegetation in complex urban scenes. To accomplish this task, we employ Mask2Former, a state-of-the-art universal image segmentation model developed by Cheng et al. (2022). This section outlines its theoretical underpinnings, architecture, and application to our specific use case.

Mask2Former: Advancing Semantic Segmentation

At its core, Mask2Former is a mask classification architecture that uses a fixed number (N) of "object queries," which can be thought of as learnable templates or detectors. Each object query, represented as a C-dimensional feature vector, starts randomly but becomes specialized through training. Processed by a Transformer decoder, these queries interact with input image features and learn to detect various objects or object parts without being pre-assigned specific types. The model uses the ground truth category labels to learn how to map the refined object queries to specific categories.

The key innovation here is that instead of using predefined anchor boxes or region proposals, the model learns to use these queries flexibly. Some queries might specialize in finding large objects, others in small objects, and some in specific shapes or textures. This flexibility enables the model to handle varying numbers and types of objects in different images, all with the same fixed set of queries.

During inference, Mask2Former processes these N object queries to predict N binary masks, each indicating which pixels belong to the respective object it has "found", along with N corresponding category labels. The model always outputs N predictions, even if there are fewer actual objects in the image. It learns to produce 'no object' predictions for unused queries, effectively ignoring them in the final output.

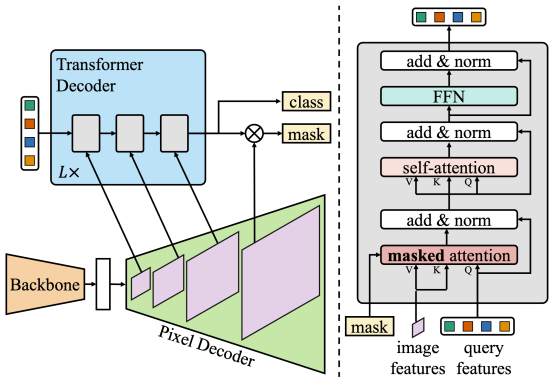

Figure 5.2: Mask2Former Architecture. Source: Cheng et al. (2022).

The architecture of Mask2Former, as illustrated in Figure 5.2, consists of three main components:

1. **Backbone**: The backbone of a neural network can be thought of as its core feature extractor. In our implementation, we utilize the SWIN-Small (Shifted Window Transformer) backbone (Liu et al., 2021). SWIN belongs to the class of Vision Transformers (ViTs), which have recently shown remarkable performance in various computer vision tasks. Unlike traditional Convolutional Neural Networks (CNNs), ViTs process images as a sequence of patches, allowing them to capture both local and global dependencies more effectively.

2. **Pixel Decoder**: This component progressively upsamples the low-resolution features from the backbone to generate high-resolution per-pixel embeddings. Crucially, it produces a feature pyramid with resolutions of 1/32, 1/16, and 1/8 of the original image. The Transformer decoder then processes these multi-scale features, with each layer operating on a different resolution. This approach enables Mask2Former to capture both fine-grained details and broader contextual information while balancing computational efficiency.

3. **Transformer Decoder**: This module processes image features to refine object queries and produce final segmentation masks. It incorporates a novel masked attention mechanism, which is key to the model's performance.

The masked attention mechanism in the Transformer decoder is a critical innovation in Mask2Former. It constrains the cross-attention to focus on relevant image regions, significantly improving both efficiency and accuracy. Mathematically, the process can be described as follows:

Query Feature Processing:

For layer l, the query features $X_l \in \mathbb{R}^{N \times C}$ (where N is the number of queries and C is the feature dimension) are processed as:

$$Q_l = f_Q(X_{l-1}) \in \mathbb{R}^{N \times C} \tag{5.1}$$

where f_Q is a linear transformation.

Masked Attention:

The masked attention operation is defined as:

$$X_l = \sigma(M_{l-1} + Q_l K_l^T) V_l + X_{l-1} \tag{5.2}$$

where $K_l, V_l \in \mathbb{R}^{H_l W_l \times C}$ are the image features under transformations f_K and f_V respectively. H_l and W_l are the spatial dimensions of the image features at layer l, and C is the feature dimension.

The softmax function σ normalizes the attention weights, converting raw attention scores into probabilities that sum to 1. For a vector of scores $z = [z_1, z_2, ..., z_n]$, the softmax function is defined as:

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{n} e^{z_j}} \tag{5.3}$$

The attention mask M_{l-1} at feature location (x, y) is defined as:

$$M_{l-1}(x, y) = \begin{cases} 0 & \text{if } M_{l-1}(x, y) = 1 \\ -\infty & \text{otherwise} \end{cases} \tag{5.4}$$

Setting the attention mask to $-\infty$ for non-relevant pixels effectively zeroes out their contribution in the softmax operation, forcing the model to completely ignore these areas.

Segmentation Head:

The final segmentation output S is produced by applying a series of operations to the refined features from the last Transformer decoder layer:

$$S = \sigma(f_u(f_c(X_L))) \tag{5.5}$$

where L is the total number of Transformer decoder layers, and:

- f_c represents a series of 3x3 convolutional layers with batch normalization and ReLU activation between them,
- f_u denotes upsampling operations to restore the spatial resolution to that of the original input image,
- σ is the softmax activation function applied per-pixel across C_{out} classes.

The output $S \in \mathbb{R}^{H \times W \times C_{out}}$ provides a probability distribution over C_{out} classes for each pixel in the $H \times W$ image.

This sophisticated architecture enables Mask2Former to synthesize fine-grained local details with broader contextual information, resulting in highly accurate semantic segmentation even in complex urban environments.

Urban Greenery Index Calculation

To translate the model's output into a quantifiable measure of urban greenery, we introduce the Urban Greenery Index (UGI). This index leverages the segmentation masks produced by Mask2Former, specifically focusing on the 'vegetation' and 'terrain' classes to capture a comprehensive view of urban green spaces. This approach provides a numerical representation of urban greenery and follows the method proposed by Zhang and Dong (2018), adapted for our Mask2Former output. Our UGI calculation combines vegetation and terrain categories to capture overall greenery in urban environments. This decision acknowledges that grass, classified as terrain by the model, is a significant component of urban green spaces. By including both categories, we achieve a more comprehensive assessment of urban greenery.

Formally, for an image j, let $S_j \in \mathbb{R}^{H \times W \times C_{out}}$ be the softmax output of the segmentation head. We define p_i as the class with the highest probability for pixel i:

$$p_i = \arg\max_c S_j(i, c) \tag{5.6}$$

where $S_j(i, c)$ is the probability of class c for pixel i in image j.

The Urban Greenery Index for image j is then computed as:

$$\text{UGI}_j = \frac{\sum_{i=1}^{N} \mathbf{1}(p_i \in \text{vegetation, terrain})}{N} \tag{5.7}$$

where N is the total number of pixels in the image, and $\mathbf{1}(\cdot)$ is an indicator function that equals 1 if the condition is true, and 0 otherwise.

This approach provides a straightforward yet effective measure of greenery presence in urban environments. By applying this calculation to our extensive dataset of street-view images, we generate a comprehensive representation of urban greenery distribution across Rotterdam, as depicted in Figure 4.2.

Implementation

For our specific application, we utilize a Mask2Former model pretrained on the Cityscapes dataset (Cordts et al., 2016). This approach aligns with common practices in the literature (Fu et al., 2019; Ye et al., 2019) and is particularly suitable given the similarities between Cityscapes street scenes and our street-view imagery. We specifically use the Mask2Former for Semantic Segmentation model (Model ID: 48333149_3), which is available in the model zoo.

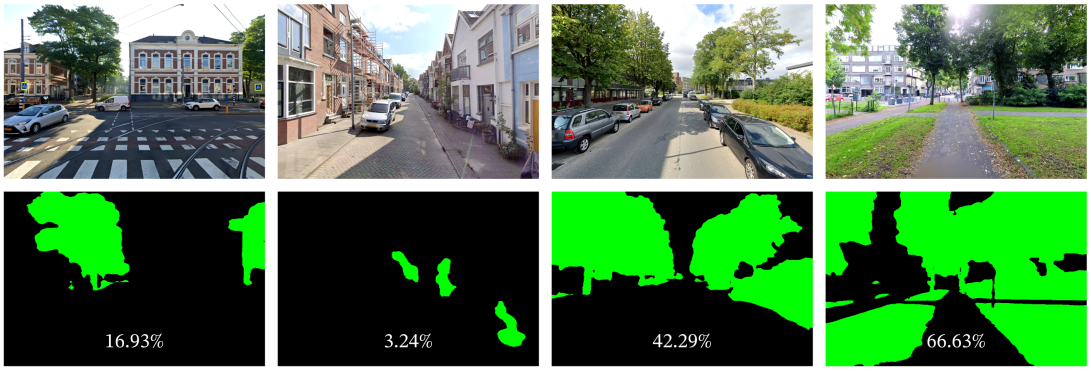

Figure 5.3: Original Images (Top) and their Segmented Masks (Bottom) reveal the extent of greenery in different urban settings. The percentages represent the Urban Greenery Index (UGI) for each scene.

Choosing a pretrained model balances computational efficiency and data constraints. Training Mask2Former from scratch demands significant computational power and a vast, annotated dataset of urban street scenes. Leveraging a pretrained model offers state-of-the-art performance without extensive fine-tuning or manual dataset annotation. Mask2Former's performance on the Cityscapes test set achieves a Mean Intersection-Over-Union (Mean IoU) of 84.3% (Cheng et al., 2022), representing the current state-of-the-art for semantic segmentation on this benchmark. The Mean IoU metric, widely used in the field, measures the overlap between predicted and ground truth segmentation masks, providing a robust indication of segmentation accuracy. Ground truth segmentation masks are accurately labeled images created by humans to serve as references for evaluating the performance of segmentation models.

While direct evaluation of the model's performance on our street-view images isn't feasible without ground truth annotations, our visual inspection of the segmentation results indicates strong accuracy in detecting vegetation. This qualitative assessment, though less rigorous than quantitative methods, is commonly used in comparable studies (Fu et al., 2019; Ye et al., 2019), especially with novel, unannotated datasets. Figure 5.3 shows example images alongside their corresponding masks.

The computational resources required for this analysis are significant. Using an M1 Pro processor and leveraging PyTorch's Metal Performance Shaders (MPS) backend for GPU training acceleration and multithreading capabilities in Python where possible, the Mask2Former inference process took approximately 140 hours (~6 days), equating to roughly 2.2 seconds of processing time per image.

Urban Activity Metrics: Detection of People, Cars, and Bicycles

To complement our analysis of urban greenery, we employ the YOLO (You Only Look Once) object detection model to quantify the presence of people, cars, and bicycles in our street-view images. YOLO, introduced by Redmon et al. (2016) and further developed in YOLO-v3 by Redmon and Farhadi (2018), which is the version we will be using, represents a paradigm shift in object detection. Unlike traditional methods that repurpose classifiers for detection, YOLO frames object detection as a single regression problem, enabling end-to-end training and real-time performance. This unified approach

allows YOLO to reason globally about the image, making it particularly suitable for analyzing complex urban scenes.

YOLO Architecture and Mathematical Framework

The YOLO model divides the input image into an $S \times S$ grid. Each grid cell is responsible for detecting objects whose center falls within its boundaries. This grid-based approach allows YOLO to process the entire image in a single forward pass, contributing to its speed and efficiency.

While the specific objects in an image do not need to be defined beforehand, the classes of objects that YOLO can detect are defined during the training phase. The model is trained on a dataset containing labeled objects, and it learns to recognize and detect only those classes present in the training set. Consequently, during inference, the model's ability to identify objects is contingent upon the classes included in its training dataset.

For each grid cell (i,j), YOLO predicts B bounding boxes and C conditional class probabilities. Each bounding box prediction consists of five components:

$$B_{i,j,b} = (x, y, w, h, \text{conf}_{i,j,b}), \quad b \in 1, ..., B \tag{5.8}$$

Here, (x, y) are the center coordinates of the box, normalized to be offsets within the grid cell (between 0 and 1), while w and h represent the width and height of the box relative to the whole image. The confidence score $\text{conf}_{i,j,b}$ reflects the model's estimate of how likely it is that the box contains an object.

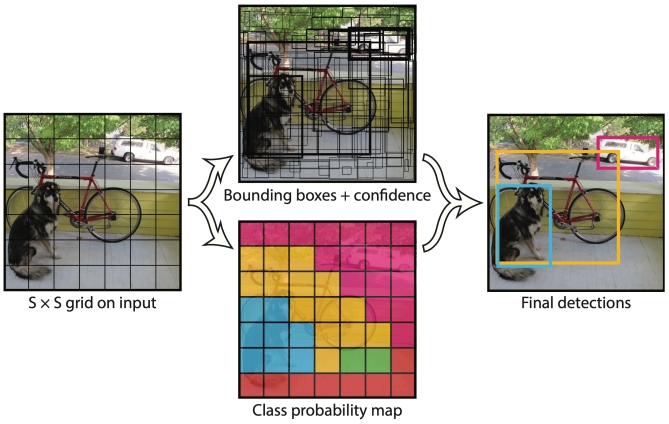

Figure 5.4: YOLO Model Framework. Source: Redmon et al. (2016).

Alongside these bounding boxes, YOLO predicts a set of conditional class probabilities for each grid cell:

$$P_{i,j,c} = P(\text{class}_c|\text{Object}), \quad c \in 1, ..., C \tag{5.9}$$

These probabilities represent the likelihood of the object belonging to each class, given that an object is present.

The interaction between bounding boxes and class probabilities is key to YOLO's functionality. While each grid cell predicts B bounding boxes, it only predicts one set of class probabilities, assuming each grid cell will contain at most one object. During inference, the confidence score $conf_{i,j,b}$ for each bounding box can be interpreted as $P(\text{Object})$, representing the model's estimate of the likelihood of an object's presence.

To obtain the final class-specific confidence scores for each box, YOLO multiplies the conditional class probabilities by the individual box confidence predictions:

$$\text{score}_{i,j,b,c} = P_{i,j,c} * \text{conf}_{i,j,b} \tag{5.10}$$

This computation yields a confidence score for each class c in each bounding box b of grid cell (i, j), effectively indicating how confident the model is that a particular box contains an object of a specific class. This detection process is visualized in Figure 5.4.

The model then eliminates low-scoring boxes and applies non-maximum suppression to remove duplicate detections. While these steps allow YOLO to efficiently predict multiple objects in various locations within the image, further specifics of this process are beyond the scope of this paper.

Model Network

The YOLO-v3 architecture, as illustrated in Supplementary Figure S8.1, is built upon a deep convolutional neural network called Darknet-53 (Redmon and Farhadi, 2018). This backbone network comprises 53 convolutional layers, strategically employing a combination of 3×3 and 1×1 convolutions. The network is organized into five main stages, each downsampling the input by a factor of 2, resulting in feature maps at different scales. YOLO-v3 capitalizes on this multi-scale feature hierarchy by making predictions at three different scales: 32×32, 16×16, and 8×8 grid cells for the input image. This multi-scale approach allows the network to detect objects across a wide range of sizes, with smaller objects benefiting from the higher resolution feature maps earlier in the network.

Furthermore, YOLO-v3 introduces several improvements over its predecessors, enabling the detection of overlapping objects, as well as objects that can belong to multiple categories simultaneously. This architecture, with its deep feature extraction and multi-scale predictions, and advanced training techniques, enables YOLO-v3 to achieve high accuracy and real-time performance in object detection tasks.

Object Counting Methodology

To analyze urban environments, we count specific objects such as people, cars, and bicycles using the output from YOLO. We ensure high-quality detections by applying a confidence score threshold $\tau = 0.95$. The final count for each class c is calculated as follows:

$$\text{Count}(\text{class}_c) = \sum_{i=1}^{|D|} \mathbf{1}(d_i = \text{class}_c) \cdot \mathbf{1}(\text{score}(d_i) > \tau) \tag{5.11}$$

where:

- $\mathbf{1}(\cdot)$ is the indicator function, equal to 1 if the condition is true and 0 otherwise.
- d_i represents a detected object in the set of detections D.
- $\text{score}(d_i)$ is the confidence score for detection d_i.

The robustness and efficacy of the YOLO model in detecting various urban object types are well-supported by recent studies, such as Naik et al. (2021), which demonstrated impressive accuracy and sensitivity in urban traffic video surveillance. Their study, which included the same object classes used here, achieved an average precision of 98.32% in detecting multiple object categories, validating the model's effectiveness in similar applications.

Implementation

In our implementation, we utilize a YOLO-v3 model pre-trained on the COCO (Common Objects in Context) dataset (Lin et al., 2014). The COCO dataset, developed by Microsoft, is particularly suitable for our urban analysis due to its diverse collection of objects in everyday scenes, including the categories of interest in our study.

Figure 5.5: Visual representation of our Object Detection results, showcasing bounding boxes highlighting People (yellow), Cars (red), and Bicycles (blue).

The choice of a pre-trained model, which was used as-is and was not fine-tuned or updated during our application, offers several advantages. Firstly, it circumvents the need for a large, annotated dataset of street-view images, which would be resource-intensive to create. Secondly, it leverages the robust performance of YOLO-v3 on the COCO dataset, which has been extensively validated. The COCO dataset's focus on common objects in context aligns well with our task of detecting people, cars, and bicycles in urban environments.

While we cannot perform quantitative evaluation on our specific street-view dataset due to the lack of ground truth annotations, visual inspection of the detection results indicates high accuracy in identifying and localizing the objects of interest Figure 5.5. This qualitative assessment approach, though less rigorous than quantitative metrics, is common in studies working with novel, unannotated datasets (e.g., Garrido-Valenzuela et al., 2023; Xu et al., 2022).

The computational efficiency of YOLO is particularly beneficial for our study, as it allows for the rapid processing of large volumes of street-view images. YOLO achieves high frames per second (fps) during detection, outperforming other models in terms of speed while maintaining comparable or superior accuracy (Redmon et al., 2016). Although our study does not require real-time object detection since we work with static images rather than video, the model's ability to process data quickly and efficiently is essential due to our reliance on consumer-grade hardware and large-scale image analysis needs.

The YOLO inference process, conducted on the same hardware setup (M1 Pro processor with PyTorch's Metal Performance Shaders (MPS) backend for GPU training acceleration and multithreading in Python), required approximately 64 hours (~2.67 days) to process our dataset of 229,980 images. This translates to roughly 1 second of processing time per image, demonstrating the computational efficiency of the YOLO architecture for our large-scale urban analysis.

5.3 Spatial Integration of Street View Data with Property Listings

To investigate the impact of urban environmental factors on property prices, we must establish a method to associate the features extracted from street-view images with individual property listings. This section outlines our spatial interpolation approach, which defines the neighborhood context for each property.

Our method involves defining circular neighborhoods of varying radii around each property listing and aggregating the features from all images within these areas. We investigate three neighborhood sizes with radii of 250m, 500m, and 1000m (illustrated in Figure 5.6). These radii were chosen to capture a range of spatial scales, from the immediate vicinity of a property to a broader neighborhood context.

While we developed this specific implementation ourselves, our method draws inspiration from and extends concepts found in related studies. The use of multiple spatial scales in our approach was inspired by Bency et al. (2017); although their study focused on satellite imagery and used multiple zoom levels to capture different scales, we applied this multi-scale concept to street-view images using varying radii. Our definition of circular neighborhoods aligns with the method used by Fu et al. (2019). However, where they employed a single set radius, we extend this concept by incorporating multiple radii. This approach not only captures a more comprehensive range of neighborhood effects but also facilitates meaningful comparisons across different spatial extents.

The spatial interpolation process can be formally described as follows:

1. Let $L = l_1, ..., l_N$ be the set of N property listings, where each l_i is associated with coordinates (x_i, y_i).
2. Let $P = p_1, ..., p_M$ be the set of M panoids (street-view image locations), each with coordinates (x_j, y_j) and associated feature vector $f_j = (UGI_j, cars_j, person_j, bicycle_j)$ representing the greenery index, car count, people count, and bicycle count, respectively.
3. For each listing l_i and radius $r \in 100, 250, 500, 1000$, we define the set of nearby panoids $P_{i,r}$ as:

$$P_{i,r} = p_j \in P : \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq r \tag{5.12}$$

4. We then compute the aggregated features for listing l_i at radius r as:

$$F_{i,r} = \frac{1}{|P_{i,r}|} \sum_{p_j \in P_{i,r}} f_j \tag{5.13}$$

where $|P_{i,r}|$ is the number of panoids within radius r of listing l_i.

5. The resulting $F_{i,r}$ is a vector containing the average greenery index, car count, people count, and bicycle count for the neighborhood defined by radius r around listing l_i.

To efficiently perform these spatial queries, we employ a k-d tree data structure, which allows for fast nearest-neighbor searches. The k-d tree is constructed using the coordinates of the panoids, enabling quick identification of all panoids within a given radius of each property listing.

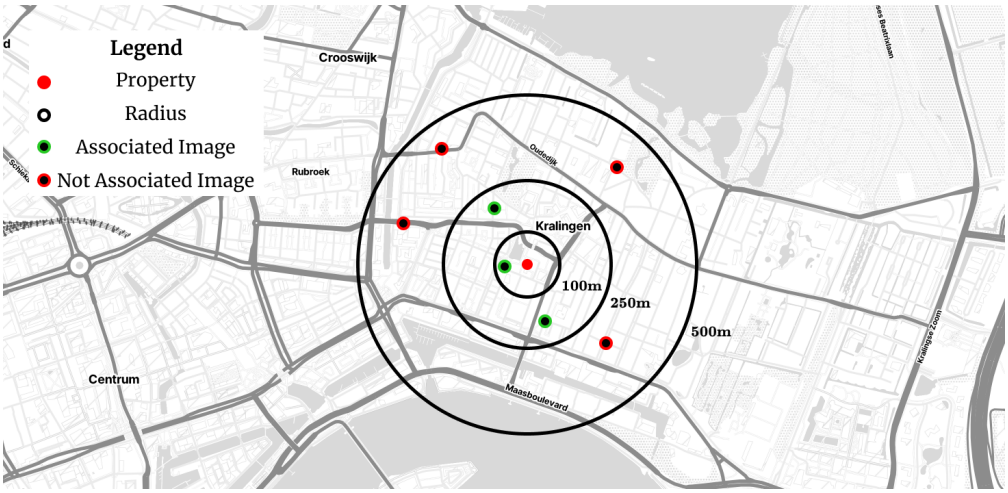

Figure 5.6: Stylized visualization of the Spatial Association (with a chosen radius of 250m).

Our approach differs from some previous studies in its flexibility and comprehensive treatment of neighborhood sizes. For instance, Fu et al. (2019) used a single fixed radius of 800m based on estimated walking ranges, while we test multiple radii to capture varying scales of neighborhood influence. In contrast to Bency et al. (2017), who used satellite imagery at different zoom levels, our method leverages street-level imagery to capture more granular, ground-level environmental features.

In contrast to our multi-scale analysis, Fu et al. (2019) utilized a single fixed radius of 800m, based on estimated walking ranges, to define their neighborhoods. Zhang and Dong (2018) similarly used a fixed radius of 400m, again based on qualitative judgment and experience, rather than empirical testing of multiple radii.

Bency et al. (2017) investigated neighborhood size using satellite images at multiple zoom levels, corresponding to increasingly larger neighborhoods centered around properties. Their findings indicated that larger neighborhood contexts provided better predictions of property prices. However, the interpolation and neighborhood definitions in their study differed significantly from ours, as they simply adjusted the zoom levels of the satellite images.

An et al. (2023) took a different approach by using a predefined number of associated images (e.g., 100, 200) to define neighborhoods. This method, while distinct, does not provide a spatially continuous measure of neighborhood context and can be less precise, particularly when considering our hypothesis that the impact of environmental factors decreases with distance from the property.

It's important to note that as the radius increases, the number of associated images per property listing also increases. This has implications for our analysis, particularly in relation to our hypothesis that the impact of environmental factors decreases with distance from the property. While larger radii provide more comprehensive neighborhood coverage, they also potentially dilute the influence of immediate surroundings. Conversely, smaller radii offer a more focused view of the property's immediate environment but may be more susceptible to local anomalies or unrepresentative samples.

Nonetheless, using different radii allows us to comprehensively investigate how the influence of environmental factors varies with distance. This multi-scale approach helps us identify the most relevant neighborhood size for different environmental features. In the next section, we will discuss our modeling approaches, including both linear and non-linear methods, to analyze the impact of these environmental features on property prices.

5.4 Modeling Framework

Linear Hedonic Pricing Model: Baseline Approach

As mentioned earlier, the hedonic pricing model posits that the price of a heterogeneous good can be decomposed into the implicit prices of its constituent characteristics (Rosen, 1974). This approach allows us to estimate the marginal contribution of various property attributes to its overall value. We begin with a standard linear hedonic pricing model as our baseline, which we then augment with visual features extracted from street-level imagery.

The general form of our hedonic pricing model can be expressed as:

$$\ln(P_i) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \epsilon_i \tag{5.14}$$

where:

- $\ln(P_i)$ is the natural logarithm of the price of property i,
- X_{ik} represents the k-th characteristic of property i,
- β_k is the corresponding coefficient and
- ϵ_i is the error term.

The use of the semi-logarithmic functional form is common in hedonic price models (Malpezzi et al., 2003; Sirmans et al., 2005) as it allows for a non-linear relationship between price and housing characteristics while maintaining linearity in parameters. This specification also provides a convenient interpretation of the coefficients: a one-unit change in X_k is associated with an approximate $\beta_k \times 100\%$ change in price, ceteris paribus.

Our baseline model incorporates standard structural and locational attributes, which we denote collectively as $\mathbf{X}$. To investigate our hypothesis that the inclusion of image-derived environmental factors significantly improves model performance compared to the baseline model with, we extend it to include our image-derived features $\mathbf{F}$:

$$\ln(P_i) = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \sum_{m=1}^{M} \gamma_m F_{im,r} + \epsilon_i \tag{5.15}$$

Here, $F_{im,r}$ represents the m-th environmental feature for property i, calculated within radius r. The coefficient γ_m captures the marginal effect of these environmental factors on the log-price. To investigate how the influence of environmental factors varies with distance, we estimate separate models for each radius r and compare their performance and coefficient estimates.

In vector notation, this extended model can be written as:

$$\ln(\mathbf{P}) = \mathbf{X}\beta + \mathbf{F}_r\gamma + \epsilon \tag{5.16}$$

where:

- $\ln(\mathbf{P})$ is an $N \times 1$ vector of the natural logarithms of property prices,
- X is an $N \times K$ matrix of property characteristics,
- β is an $K \times 1$ vector of coefficients,
- $\mathbf{F}_r$ is an $N \times M$ matrix of environmental features for each property within radius r,
- γ is an $M \times 1$ vector of coefficients for the environmental features,
- ϵ is an $N \times 1$ vector of error terms.

The Ordinary Least Squares (OLS) method is employed to estimate these models, under the standard assumptions of linearity, independence, homoscedasticity, and normality of residuals. While OLS provides easily interpretable results and has been widely used in hedonic pricing literature, it's important to note its limitations. In particular, the assumption of linearity may be overly restrictive, as Rosen (1974) already pointed out that the true relationship between housing characteristics and prices is likely to be non-linear. Moreover, the potential presence of multicollinearity among predictor variables, especially when incorporating multiple environmental factors, may affect the stability and interpretability of our estimates. To address this, we conduct Variance Inflation Factor (VIF) analysis to assess the degree of multicollinearity in our models.

While this linear hedonic model provides a solid foundation for our analysis, it may not capture complex, non-linear relationships between environmental factors and housing prices. Furthermore, potential interaction effects between variables might be overlooked. To address these limitations and test our hypothesis regarding the superior predictive performance of non-linear models, we also employ a Random Forest model, which will be introduced in the next section.

Random Forest: Capturing Non-Linear Relationships

To address the potential non-linearities and complex interactions in housing price determinants, we employ the Random Forest (RF) model as a non-parametric alternative to the linear hedonic pricing model. Introduced by Breiman (2001), Random Forest is an ensemble learning method that combines multiple decision trees to improve predictive performance and control overfitting. Several studies have utilized the Random Forest model for property valuation, finding it to outperform traditional OLS models. Hong et al. (2020) and Potrawa and Tetereva (2022) demonstrated the superior predictive accuracy of Random Forests in this context, highlighting their capability to capture complex interactions and non-linearities in housing data.

The Random Forest model can be formalized as follows:

Let $\{\mathcal{T}_b\}_{b=1}^{B}$ represent the ensemble of B decision trees in the Random Forest. For a given property i with features $\mathbf{Z}_i = [\mathbf{X}_i, \mathbf{F}_{i,r}]$, where $\mathbf{X}_i$ are the structural and locational attributes and $\mathbf{F}_{i,r}$ are the image-derived features at radius r, the prediction from the b-th tree is denoted as $\hat{P}_i^{(b)}(\mathbf{Z}_i)$. The final prediction of the Random Forest is the average of all individual tree predictions:

$$\hat{P}_i = \frac{1}{B} \sum_{b=1}^{B} \hat{P}_i^{(b)}(\mathbf{Z}_i) \tag{5.17}$$

Each tree in the forest is constructed using the following procedure:

1. **Bootstrap Sampling**: A bootstrap sample of size N is drawn with replacement from the training data.
2. **Tree Growing**: At each node:
    - Randomly select $m \ll K$ features from the full set of K features.
    - Choose the best split among these m features based on the mean squared error criterion.
    - Split the node into two child nodes.
3. **Maximization**: The tree is grown to its maximum size, typically until a minimum node size is reached.

This process is repeated B times to create the forest. The randomness introduced in both the bootstrap sampling and feature selection at each split contributes to the model's robustness and generalization ability.

One of the key advantages of Random Forest in our context is its ability to capture non-linear relationships and interactions between variables without explicit specification. This is particularly valuable for modeling the complex interplay between environmental factors and housing prices, which may not follow simple linear patterns.

Implementation

To ensure a fair comparison with the linear hedonic model, we use the same set of features $\mathbf{X}$ and $\mathbf{F}_r$ in our Random Forest models. The implementation of our Random Forest models follow these steps:

1. **Data Splitting**: The dataset is randomly divided into training (80%) and testing (20%) sets. This split is consistent across all models.
2. **Cross-Validation**: We employ 5-fold cross-validation on the training set to tune hyperparameters and assess model stability.
3. **Hyperparameter Tuning**: An exhaustive grid search is conducted to optimize key hyperparameters, including the number of trees (B), the maximum depth of each tree, and the number of features m considered for splitting at each node.

By employing both linear hedonic and Random Forest models, we are able to leverage the interpretability of linear models while capturing potential non-linearities and interactions through Random Forests, thereby addressing our research questions from multiple perspectives.

To realize the benefits of this dual approach, careful evaluation and interpretation of both models is critical. In the following sections, we will discuss the interpretation techniques used to assess and compare the performance of our linear and non-linear models.

Model Interpretation Techniques

Feature Importance

Feature importance provides a measure of each variable's contribution to the model's predictions. For Random Forests, we use the impurity-based feature importance, also known as Mean Decrease in Impurity (MDI). This method calculates the total decrease in node impurity (measured by Mean Squared Error for regression) averaged over all trees of the forest. Impurity refers to the uncertainty or disorder in the data at a particular node. When training a decision tree within a Random Forest, the model aims to minimize this impurity at each node by making splits based on different features.

For a feature j, its importance is calculated as:

$$\text{Importance}(j) = \frac{1}{B} \sum_{b=1}^{B} \sum_{t \in \mathcal{T}_b} \Delta\text{MSE}_t(j) \tag{5.18}$$

where:

- B is the number of trees in the forest
- $\mathcal{T}_b$ is the set of nodes in tree b
- $\Delta\text{MSE}_t(j)$ is the decrease in mean squared error resulting from splits on feature j in node t.

This measure quantifies how much each feature contributes to reducing the prediction error across all trees in the forest. Features that result in larger decreases in MSE are considered more important, as they play a more significant role in improving the model's predictions.

Partial Dependence Plots

Partial Dependence Plots (PDPs) illustrate the marginal effect of a feature on the predicted outcome of a machine learning model. They aid in visualizing and understanding the relationship between individual features and the model's predictions, especially in capturing non-linearities and interactions.

For a feature X_s, the partial dependence function is defined as:

$$\text{PD}_s(X_s) = \mathbb{E}_{X_C}[f(X_s, X_C)] \tag{5.19}$$

where X_C represents all other features, and f is the machine learning model's prediction function. This function gives us the expected value of f when varying X_s, averaging over all possible values of X_C.

In practice, this expectation is approximated using the training dataset:

$$\text{PD}_s(X_s) \approx \frac{1}{n}\sum_{i=1}^{n} f(X_s, x_{iC}) \tag{5.20}$$

Here, x_{iC} denotes the values of all features except X_s for the i-th instance in the dataset, and n is the number of instances.

PDPs are effective tools for understanding how changing X_s affects the model's predictions, showcasing whether the relationship is linear, non-linear, monotonic, or more complex.

SHAP (SHapley Additive exPlanations) Values

SHAP values, introduced by Lundberg and Lee (2017), provide a sophisticated method for understanding feature importance and interpreting machine learning models, leveraging principles from cooperative game theory. They offer both global insights into feature relevance across the dataset and local explanations for individual predictions, making them invaluable for understanding complex model behaviors.

For an instance x, the SHAP value for feature j is calculated as:

$$\phi_j(f, x) = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f_x(S \cup \{j\}) - f_x(S)] \tag{5.21}$$

where:

- N represents the set of all features.
- S is any subset of features excluding j.
- $f_x(S)$ is the model's prediction when using only the features in S.

To better understand SHAP values, they offer:

- **Local Interpretation**: For a specific prediction x, SHAP values quantify how much each feature contributes to the difference between the actual prediction $f(x)$ and the model's average prediction.
- **Global Interpretation**: By aggregating SHAP values across all instances in the dataset, we gain insights into the overall impact of each feature on model predictions.

Key advantages of SHAP values include their ability to:

- **Capture Feature Interaction**: They account for interactions between features, providing a nuanced view of how combinations of features influence predictions.
- **Model Non-linearity**: SHAP values capture non-linear relationships between features and predictions, offering insights into complex model behaviors that linear methods may overlook.
- **Theoretical Foundation**: Grounded in game theory, SHAP values ensure consistent and fair attribution of feature importance across different contexts and datasets.

In the following Analysis and Results section, we will apply these evaluation metrics and interpretation techniques to our models, presenting a comprehensive assessment of how urban environmental factors influence property values in Rotterdam. We will compare the performance of the linear hedonic and Random Forest models, analyze the importance and effects of different environmental features, and discuss the implications of our findings for urban planning and real estate valuation.

6 Results

This chapter presents the empirical findings of our study on the impact of urban environmental factors, extracted from street-level imagery, on residential property prices in Rotterdam. Our analysis is structured to address the four main hypotheses proposed earlier: The inclusion of urban environmental factors will enhance housing price predictions; specific environmental factors will have directional effects on housing prices, potentially differing between houses and apartments; the impact of environmental factors will depend on the distance from a property that they are being studied at; and the relationships between environmental factors and housing prices will exhibit non-linear patterns and threshold effects.

To comprehensively examine these hypotheses, we employ a multi-faceted modeling approach. Our analysis compares two distinct model types: Ordinary Least Squares (OLS) regression, representing a traditional linear approach, and Random Forest (RF), a non-linear machine learning method. This comparison allows us to assess whether the complex relationships between environmental factors and property prices are better captured by more flexible models.

Furthermore, we investigate the spatial scale of environmental effects by analyzing data at different radii around each property. This approach enables us to determine the most relevant neighborhood size for different environmental features and to test our proximity effects hypothesis.

Lastly, we conduct separate analyses for the entire dataset, as well as for houses and apartments individually. This stratification helps us uncover potential differences in how environmental factors influence property values across different property types.

Distribution of Urban Environmental Features

Before delving into our model results, it's crucial to understand the spatial distribution of the key environmental features across Rotterdam. Figure 6.1 provides a visual representation of how our image-derived neighborhood features are distributed across the city of Rotterdam.

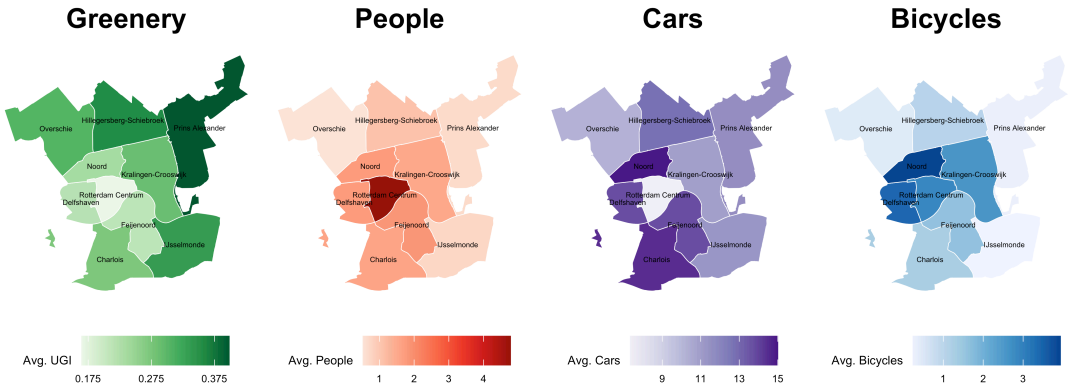

Figure 6.1: Distribution of Urban Environmental Features across Rotterdam.

Greenery (UGI): The map reveals a clear pattern of green space distribution. Rotterdam Centrum exhibits notably low levels of greenery, which is typical of dense urban cores. As we move outward, particularly towards the north, the presence of greenery increases significantly.

People: The distribution of people captured in street-view images closely aligns with urban density patterns. The city center shows the highest concentration, gradually decreasing as we move towards the periphery. This pattern likely reflects the bustling nature of Rotterdam's central business and commercial districts, contrasting with the quieter residential areas surrounding them.

Cars: Interestingly, the car distribution presents almost an inverse pattern to that of people. While one might expect a high concentration in the city center, we observe fewer cars in central areas, possibly due to pedestrianized zones and public transport efficiency. Notably, Rotterdam Noord shows a surprisingly high car presence, which could be indicative of commuter patterns or parking policies in these areas.

Bicycles: The bicycle distribution partially mirrors the people distribution, with concentrations in central and mixed-use areas. However, suburban regions show lower bicycle presence, likely due to a combination of factors: lower population density, increased private storage options (garages, sheds), and potentially different commuting habits.

These distributions provide valuable context for interpreting our subsequent analyses, highlighting the spatial heterogeneity of urban environmental features across Rotterdam. They underscore the importance of considering location-specific factors when analyzing property prices and urban dynamics.

Throughout this chapter, we present our results using a combination of tables and figures, evaluating model performance through metrics such as R-squared, Mean Squared Error (MSE), and Mean Absolute Percentage Error (MAPE). We also employ visualization techniques including partial dependence plots and SHAP values to interpret the complex relationships uncovered by our models.

6.1 Comparative Analysis of Model Performance

We begin our analysis by examining the overall performance of our models and assessing the impact of incorporating image-derived environmental features. For this initial

overview, we focus on models using the large 1000m radius, trained on 80% of the data (5319 observations), with performance metrics (R^2 and MSE) calculated on the held-out test set (20% , 1330 observations).

Table 6.1 presents a comparison of OLS and RF models, showing their performance with different combinations of features. The baseline models include only structural attributes, while subsequent models incrementally incorporate individual image features (UGI, People, Cars, Bicycles) before combining all features. To ensure the validity of our models, we assessed multicollinearity among predictors using Variance Inflation Factors (VIF), the results of which can be found in Supplementary Figure S8.2.

Table 6.1: Comparison of OLS and RF models across various features.

	Baseline		UGI		People		Cars		Bicycles		All	
Model	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE	R^2	MSE
OLS	0.725	1.000	0.726	1.032	0.753	1.077	0.733	**0.963**	0.757	1.131	**0.775**	1.021
RF	0.788	0.695	0.824	0.570	0.834	0.585	0.818	0.604	0.842	0.592	**0.905**	**0.349**

Note:
MSE values are relative to the OLS Baseline, which is set to 1. Lower is better.

Examining the baseline models, which include only structural attributes, we observe that the RF model ($R^2 = 0.788$, MSE = 0.695) outperforms the OLS model ($R^2 = 0.725$, MSE = 1). This initial difference suggests that even with traditional housing characteristics, the RF model is better at capturing the patterns present in the data. Notably, the MSE for the OLS baseline model is set at 1, with all other values presented relative to this baseline.

As we incorporate image features, the performance disparity between OLS and RF models becomes more pronounced. For RF models, each image feature contributes to improved performance. The Urban Greenery Index (UGI) provides the largest individual improvement, increasing R^2 from 0.788 to 0.824 and decreasing MSE from 0.695 to 0.570. This substantial improvement underscores the importance of green spaces in determining property values, aligning with findings from previous studies (Ye et al., 2019; Zhang and Dong, 2018).

Conversely, OLS models show only marginal improvements with the addition of individual image features. The inclusion of the Cars variable yields a slight MSE improvement (from 1.000 to 0.963), while other features show minimal impact. While R^2 values for OLS models do increase slightly with each feature addition, this suggests that linear models may struggle to capture the complex relationships between these environmental factors and property prices.

The most striking contrast between OLS and RF performance is evident when all image features are included. The RF model with all features demonstrates the best overall performance ($R^2 = 0.905$, MSE = 0.349), representing a substantial improvement over the baseline. In contrast, the OLS model with all features shows a more modest improvement in R^2 (0.775) and a slightly worse MSE (1.021) compared to its baseline. This discrepancy suggests that OLS models may struggle to effectively incorporate the complex information provided by image features, possibly due to interactions that the linear model cannot capture.

The relative importance of different image features also varies between model types. For RF models, UGI appears to be the most influential individual feature, followed by

Bicycles, People, and Cars. However, the best performance is achieved when all features are combined, indicating that these environmental factors provide complementary information about property values.

These findings support our first hypothesis regarding enhanced predictive accuracy, particularly for RF models. The substantial improvements in both R^2 and MSE when incorporating image features demonstrate the value of including urban environmental factors in property valuation models. However, this hypothesis is only partially supported for OLS models, which show limited ability to leverage the additional information effectively.

The significant performance difference between OLS and RF models when incorporating image features has important implications for real estate valuation. It shows that the relationships between visual environmental factors and property prices are likely complex, requiring more flexible models to capture these dynamics accurately, aligning with the findings by Hong et al. (2020). In the following section, we will explore the specific directional effects of each environmental factor on property prices, examining how these impacts differ between houses and apartments.

6.2 Directional Impacts of Environmental Features

Building upon our initial model performance analysis, we now delve into the specific directional effects of environmental features on housing prices, which addresses our second hypothesis.

To investigate these effects, we estimated Ordinary Least Squares (OLS) models for the entire dataset (N = 6,649), as well as separate models for apartments (N = 5,001) and houses (N = 1,648). All models use a 1000m radius for environmental features, with coefficients and other statistics calculated on the full respective datasets. Table 6.2 presents the coefficients for different house types, while Figure 6.2 visualizes the standardized coefficients of image-derived features for direct comparability.

Table 6.2: OLS Coefficients across House Types.

	All Data		Apartments Only		Houses Only	
Variable	Coefficient	95% CI	Coefficient	95% CI	Coefficient	95% CI
(Intercept)	10.657***	(10.247, 11.068)	10.310***	(9.866, 10.754)	11.503***	(10.620, 12.386)
Living Area (m²)	0.009***	(0.008, 0.009)	0.009***	(0.009, 0.010)	0.007***	(0.007, 0.008)
Energy Label						
Higher (A+, A, B, C)	—	—	—	—	—	—
Lower (D, E, F, G)	-0.050***	(-0.062, -0.038)	-0.060***	(-0.072, -0.047)	0.043**	(0.011, 0.074)
Rooms	-0.012***	(-0.018, -0.005)	-0.037***	(-0.044, -0.029)	0.010	(-0.003, 0.022)
Construction Year	0.001***	(0.000, 0.001)	0.001***	(0.001, 0.001)	0.000	(0.000, 0.001)
Insulation						
Double Glass	—	—	—	—	—	—
Fully Insulated	0.171***	(0.155, 0.187)	0.165***	(0.148, 0.183)	0.174***	(0.139, 0.209)
Other	0.007	(-0.006, 0.021)	0.008	(-0.006, 0.022)	-0.019	(-0.049, 0.012)
Heating						
Central Heating	—	—	—	—	—	—
District Heating	0.044***	(0.027, 0.060)	0.053***	(0.036, 0.071)	0.015	(-0.020, 0.050)
Other	-0.047***	(-0.062, -0.032)	-0.027***	(-0.042, -0.011)	0.016	(-0.026, 0.058)
UGI	0.490***	(0.410, 0.571)	0.219***	(0.128, 0.309)	0.894***	(0.733, 1.055)
People	0.000	(-0.007, 0.008)	-0.019***	(-0.028, -0.011)	0.020	(-0.015, 0.056)
Cars	-0.021***	(-0.024, -0.018)	-0.025***	(-0.028, -0.022)	-0.008*	(-0.015, -0.002)
Bicycles	0.091***	(0.084, 0.098)	0.106***	(0.099, 0.113)	0.078***	(0.051, 0.105)
R^2	0.777		0.772		0.702	
R^2 *(adj.)*	0.776		0.771		0.700	
p-value	<0.001		<0.001		<0.001	
N	6649		5001		1648	

[1] *p<0.05;* ***p<0.01;*** p<0.001
[2] CI = Confidence Interval

Structural Attributes

Before examining our primary variables of interest, we first consider the effects of structural attributes across different property types. Most coefficients align with expectations from the literature, but several noteworthy patterns emerge:

1. **Living Area**: Positive and significant across all models, confirming then expected premium for larger properties.
2. **Energy Label**: The lower (worse) energy label category negatively impacts prices for apartments and the complete data but shows a positive coefficient for houses. This might reflect that older, less energy-efficient houses command higher prices due to their location in prime areas of the city or unique architectural features.
3. **Number of Rooms**: Negative and significant for apartments and overall, indicating a preference for fewer, larger rooms. This effect is not significant for houses.
4. **Year Built**: Significant and positive for apartments and overall, but not for houses, suggesting the age of a house may be less important. However, as Supplementary Figure S8.3 implies, the impact on house prices is U-shaped, where both very old and very new houses are valued highly.
5. **Insulation**: Full Insulation is preferred across all datasets, highlighting the value of energy efficiency.
6. **Heating**: Significant for apartments with district heating, possibly due to its convenience and cost-efficiency in densely populated areas, but not significant for

houses, which may be due to buyers' flexibility to choose or upgrade their heating systems over time, reducing the relative impact of heating type.

These variations in structural attribute effects highlight the importance of considering property type in valuation models, as buyers of houses and apartments appear to value features differently.

Image-Derived Environmental Features

Turning to our primary focus, we analyze the effects of image-derived environmental features across different property types:

1. **Urban Greenery Index (UGI):** The UGI shows a positive and significant effect across all datasets, strongly supporting our hypothesis. As expected, the effect is most pronounced for houses, with a coefficient approximately four times larger than for apartments ($\beta_{UGI}^{Houses} = 0.894$ vs. $\beta_{UGI}^{Apartments} = 0.219$). This substantial difference suggests that house buyers place a higher premium on surrounding greenery, possibly due to lifestyle preferences, such as a desire for family-friendly environments, which they value more highly compared to apartment buyers. To put this in perspective, a 0.1 increase in the UGI is associated with an approximately 9.4% increase in house prices, ceteris paribus.
2. **Presence of People:** Contrary to our initial hypothesis, the effect of People is only significant for apartments, and unexpectedly negative ($\beta_{People}^{Apartments} = -0.019$). In the context of Rotterdam, a higher concentration of people in apartment-dominated areas might be associated with issues such as noise, lack of privacy, or overcrowding. For houses, the coefficient is positive but not significant, suggesting that in less dense areas, the presence of people might indeed indicate neighborhood vitality as initially hypothesized. We will explore this variable further in the context of different radii in a subsequent section.

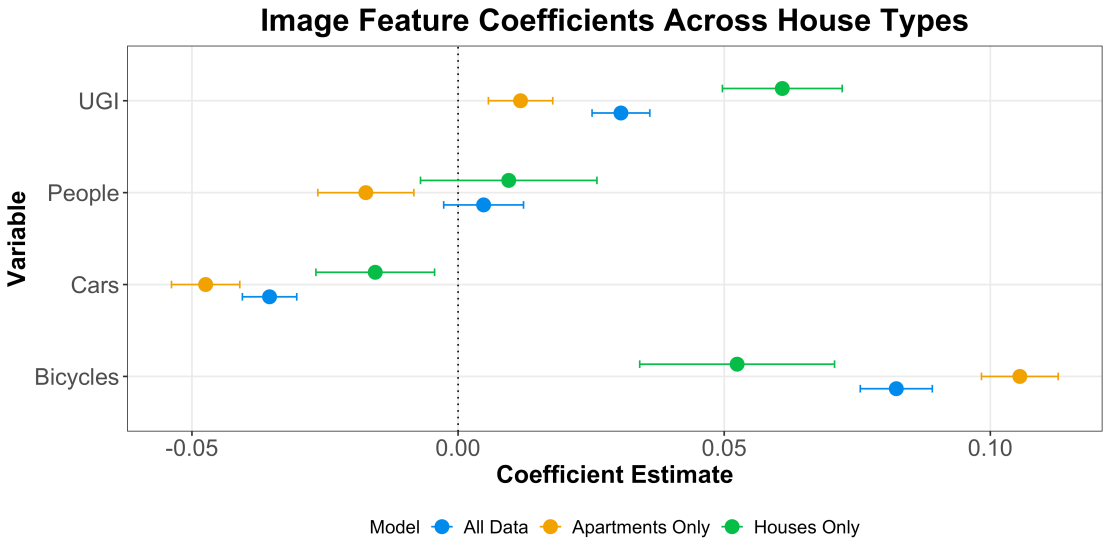

Figure 6.2: OLS Coefficients Plot across House Types.

3. **Number of Cars:** The presence of cars shows a significant negative effect across all models, supporting our hypothesis. However, contrary to our expectations, the effect is about three times stronger for apartments ($\beta_{Cars}^{Apartments} = -0.025$) compared to houses ($\beta_{Cars}^{Houses} = -0.008$). This stronger negative effect for apartments might reflect several factors:

    a) Apartment dwellers may be less likely to own cars and thus more sensitive to the negative externalities of high car presence (e.g., noise, air pollution, reduced pedestrian safety).
    b) Areas with high car presence might correlate with busy roads or commercial areas, which could be less desirable for residential apartments.
    c) Houses in car-dense areas might have offsetting benefits (e.g., better accessibility, larger plots) that partially mitigate the negative effects.

4. **Number of Bicycles:** The presence of bicycles shows a positive and significant effect across all datasets, supporting our hypothesis. As expected, the effect is more pronounced for apartments ($\beta_{Bicycles}^{Apartments} = 0.106$) compared to houses ($\beta_{Bicycles}^{Houses} = 0.078$). This aligns with our expectation that apartments, often located in more central, bicycle-reliant urban areas, would benefit more. Specifically, the presence of one more bicycle is associated with an approximately 10.6% increase in apartment prices, ceteris paribus. The positive effect likely captures several desirable neighborhood characteristics:

    a) Good cycling infrastructure, which may correlate with overall neighborhood quality and sustainability.
    b) Proximity to amenities, as areas with high bicycle use often have mixed land use and good local services.

Figure 6.2 reveals additional insights into the relative importance of these environmental features across property types. Notably, the houses-only model consistently shows the highest degree of uncertainty, as evidenced by the wider confidence intervals. This increased uncertainty stems from the smaller sample size (N = 1,648) compared to the apartments and overall datasets. Across all models, the People variable demonstrates a relatively small effect. Interestingly, while the Urban Greenery Index (UGI) exhibits the largest effect among all features for houses, it shows the smallest impact for apartments. This stark contrast validates our initial expectations. Despite these differences, the R^2 and adjusted R^2 values are similar across the three models, indicating consistent explanatory power despite the different subsets of data. The lower R^2 for houses might reflect greater heterogeneity in this subset or the influence of unobserved factors not captured by our current set of variables.

Implications

These findings highlight the complexity of how environmental features impact property values and underscore the importance of differentiating between property types in urban planning and real estate valuation.

The strong positive effect of the Urban Greenery Index (UGI), particularly for houses, highlights the critical role of green space in urban development. The availability and extent of green spaces should be a focus of urban planners, especially in areas with single-family homes. The differing effects of transportation infrastructure on houses and apartments reflect distinct transportation needs and preferences. In apartment-focused

developments, prioritizing bicycle infrastructure and reducing car presence could enhance property values. Conversely, for areas with more single-family houses, a balanced approach also accommodating cars might be more suitable.

The varying impacts of environmental features on houses versus apartments likely mirror differing buyer preferences and lifestyles. House buyers may value green spaces and quieter environments more, while apartment buyers might prioritize urban amenities and sustainable transport options. In the next section, we will examine how these effects vary across different spatial scales, focusing on the changing impact of environmental factors at various distances from the property.

6.3 Spatial Variation in Environmental Effects

Building on our analysis of directional effects, we now examine how the impact of environmental features varies across different spatial scales.

To investigate these spatial effects, we estimated both OLS and Random Forest (RF) models for three different radii: 250m, 500m, and 1000m. Table 6.3 presents the performance metrics (MAPE and R^2) for both model types across these radii, calculated on the 20% test set. Additionally, Figure 6.3 illustrates the standardized OLS coefficients for image-derived features at different radii, estimated on the full dataset to provide a clear interpretation of how these effects change with distance.

Model Performance Across Spatial Scales

Examining Table 6.3, we observe a consistent trend of improving model performance as the radius increases, for both OLS and RF models: For the former, the R^2 increases from 0.749 at 250m to 0.775 at 1000m, while the MAPE decreases from 0.172 to 0.162, while the latter RF models' R^2 rises from 0.874 at 250m to 0.905 at 1000m, with the MAPE dropping from 0.114 to 0.098.

These results suggest that broader neighborhood characteristics captured by larger radii provide more explanatory power for housing prices. This trend contradicts our initial hypothesis of decreasing influence with distance and aligns more closely with findings from Bency et al. (2017), who observed that larger neighborhood contexts often yield better predictions of property values.

Table 6.3: Comparing OLS and RF Models across Different Radii.

	OLS			RF		
Metric	250m	500m	1000m	250m	500m	1000m
MAPE	0.172	0.168	**0.162**	0.114	0.105	**0.098**
R^2	0.749	0.761	**0.775**	0.874	0.891	**0.905**

Comparing OLS and RF models, we note that RF outperforms OLS across all radii, with consistently higher R^2 and lower MAPE values. Moreover, the performance gap between RF and OLS widens as the radius increases, with a difference in R^2 of 0.125 at 250m vs. 0.130 at 1000m.

Spatial Variation of Environmental Effects

Figure 6.3 illustrates how the effects of different environmental features change across spatial scales, allowing for a clearer interpretation of spatial trends. The standardization of coefficients enables direct comparison of effect sizes across features and radii.

1. **Urban Greenery Index (UGI):** The effect of UGI becomes more pronounced and statistically significant as the radius increases. At 250m, the effect is not significant, but it becomes strongly positive at 500m and 1000m. This pattern suggests that the overall green character of a broader neighborhood is more influential on property values than immediate green spaces. This could reflect buyers' preference for generally greener areas, where they can benefit from larger parks or green corridors that might not be captured in smaller radii.
2. **Presence of People:** Interestingly, this variable shows a different pattern compared to other features. It has a significant positive effect at smaller scales but becomes insignificant at 1000m. This unique trend might indicate that the vitality and liveliness of the immediate neighborhood (as proxied by the presence of people) positively influences property values, but this effect diminishes at larger scales. Buyers might value a lively local environment but be less concerned with activity levels in the broader area.

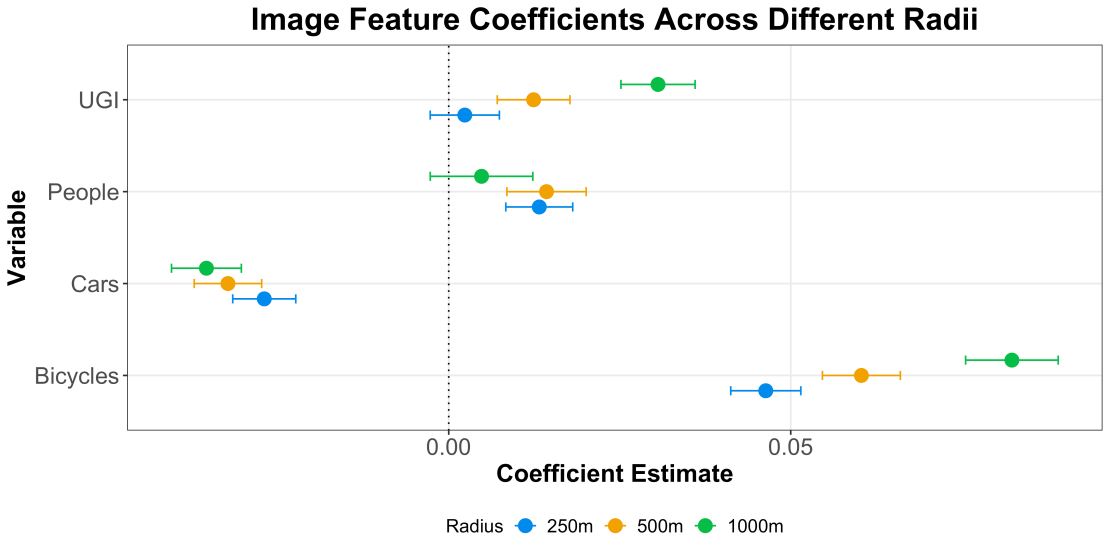

Figure 6.3: OLS Coefficients Plot across Different Radii.

3. **Number of Cars:** The negative effect of cars becomes slightly stronger as the radius increases, with the largest effect at 1000m. This could suggest that while local traffic might be a concern, the overall car presence in a broader area has a more significant impact on property values. This might reflect buyers' considerations of general congestion levels, air quality, or the car-dependency of a larger neighborhood.
4. **Number of Bicycles:** The positive effect of bicycles also intensifies with increasing radius, showing the strongest effect at 1000m. This trend validates our interpretation of bicycle presence as a proxy for cycling infrastructure. Buyers appear to value good cycling connectivity across a broader area, not just in the

immediate vicinity of a property. This aligns with the idea that the utility of extensive cycling infrastructure extends well beyond the local street level.

Controlling for Area Effects

To further investigate spatial effects and control for potential area-related confounders, we integrated the "Area" variable (defined as Rotterdam's *Wijken*, aligned with those depicted in Figure 6.1) into all model specifications outlined in Table 6.3. This analysis aimed to discern whether model enhancements were primarily driven by image-derived features or broader neighborhood characteristics.

As Table 6.4 reveals, including the "Area" variable significantly boosted OLS model performance. The baseline OLS model, incorporating area but excluding image features, outperformed models with image features but without area. This suggests that area-level factors explain more of the variance in property values than image-derived features within the OLS framework. Specifically, introducing the area variable increased the baseline model's goodness of fit by over 10 percentage points (from 0.725 to 0.829), while subsequently incorporating environmental factors yielded only marginal improvements.

Table 6.4: Comparing OLS and Random Forest Models across Different Radii and Subsets of the Data.

		OLS				RF			
Data	Metric	None	250m	500m	1000m	None	250m	500m	1000m
Area Excluded									
	MAPE	0.188	0.172	0.168	**0.162**	0.159	0.114	0.105	**0.098**
	R^2	0.725	0.749	0.761	**0.775**	0.788	0.874	0.891	**0.905**
Area Included									
	MAPE	0.139	0.139	0.138	**0.137**	0.150	0.102	0.100	**0.097**
	R^2	0.829	0.830	0.831	**0.834**	0.797	0.902	0.904	**0.908**
Area 'Noord'									
	MAPE	0.122	0.120	0.118	**0.116**	0.117	**0.098**	0.100	0.101
	R^2	0.678	0.689	0.696	**0.710**	0.705	**0.769**	0.761	0.767

Random Forest models presented a contrasting picture. In the baseline scenario without image features, adding the area variable resulted in a modest R^2 increase from 0.788 to 0.797. Introducing image features substantially improved performance across all radii, with minimal variation between radii. Regardless, the model with the 1000m radius consistently exhibited the highest performance based on both MAPE and R^2. Notably, at this radius, incorporating the area variable did not significantly enhance model performance compared to using image features alone.

Interestingly, the relationship between model performance and radius size changed when controlling for the area. In models using only image features, performance generally increased with larger radii, suggesting that these models were partially capturing area-level effects. However, when area was included as a control variable, this pattern disappeared. To further assess the specific impact of the environmental features themselves, we further conducted an analysis focusing on a specific area.

Evidence from Rotterdam Noord

To isolate the impact of environmental features and delve deeper into the spatial effects within a more homogenous context, we conducted a focused analysis on Rotterdam Noord. This *Wijk* has the largest number of properties (1021) in our dataset, providing a robust sample size for our analysis. By examining a single area, we mitigate the influence of broader area-level factors and concentrate on the impact of environmental features on property values within a more localized setting.

As the bottom section of Table 6.4 indicates, OLS models exhibited slight improvements when including environmental features, mirroring the city-wide analysis without the area variable. Performance gradually increased with larger radii.

Conversely, Random Forest models revealed a different pattern. While environmental features still significantly improved performance, the best-performing model shifted to the smallest radius (250m), contrary to the city-wide results. This suggests that within a single area, where confounding effects are minimized, environmental features at smaller radii exert a more substantial influence on property values, supporting our initial hypothesis that people prioritize their immediate surroundings over the broader neighborhood context.

Implications

Contrary to our initial hypothesis, the impact of most environmental factors appears to increase with distance rather than decrease, suggesting that broader neighborhood characteristics may be more influential than immediate surroundings for most features. The presence of people stands out as the exception, showing the strongest proximity effect.

These findings have several implications for real estate valuation and urban planning. Our performance metrics show that models consistently extract additional relevant information from most environmental features as the radius increases. Future research could also explore even larger radii to determine an upper limit to this effect. Additionally, developing models that allow for different optimal scales depending on the variable could further refine our understanding of these spatial relationships.

The focused analysis on Rotterdam Noord provides critical insights into how environmental features affect property values at a more localized level and challenges our city-wide findings. The result that smaller radii are more influential in this specific area suggests that local environmental characteristics have a substantial impact on housing prices when controlling for larger-scale area differences.

Urban planners should consider these broader impacts when designing neighborhoods or implementing infrastructure improvements. Our findings suggest that larger-scale initiatives, such as extensive cycling networks and broader green space planning, could have significant positive impacts on property values across wider areas. Conversely, in specific neighborhoods like Rotterdam Noord, enhancing local features such as greenery and pedestrian-friendly spaces might yield more immediate benefits.

There is an important caveat to our analysis at varying spatial scales. As the radius increases, the average number of images associated with each property also rises (208 at 250m, 780 at 500m, 2824 at 1000m). This increase in data points per property at larger radii has both statistical and practical implications: The larger number of images

at greater radii may contribute to more stable and precise estimates of environmental features, potentially explaining some of the improved model performance at larger scales. Future work might explore different aggregation methods, such as setting a fixed number of associated images or weighted averages based on distance, to refine the capture of spatial effects.

6.4 Uncovering Non-Linear and Threshold Effects

Our analysis thus far has revealed complex relationships between environmental features and property values, varying across housing types and spatial scales. Building on these insights, we now examine how the impact of environmental factors may shift at different levels of intensity or prevalence, aiming to uncover potential tipping points or saturation levels in their influence on housing prices.

To investigate these nuanced relationships, we employ a Random Forest model, complemented by SHAP analysis and Partial Dependence Plots. This approach allows us to capture and visualize complex, non-linear patterns that may not be apparent in linear models. We trained our RF model using 80% of the dataset, with environmental features calculated at a 1000m radius - the specification identified in the preceding section as the most accurate model configuration. SHAP values and partial dependence functions were computed using the 20% test data, though our experiments showed similar results when utilizing the training data or the entire dataset, indicating the robustness of these findings.

Variable Importance

Figure 6.4 presents the variable importance for our RF model, utilizing the impurity-based importance method. This approach measures the average reduction in node impurity (in this case, the mean squared error for our regression task) achieved by splitting on each feature across all trees in the forest. Features that consistently lead to larger reductions in impurity are considered more important, as they contribute more significantly to the model's predictive accuracy.

The plot reveals a clear hierarchy of feature importance in determining housing prices. Living area emerges as the most crucial predictor, followed by the number of rooms and construction year. This aligns with traditional real estate valuation principles, affirming the primacy of core structural attributes in determining property values.

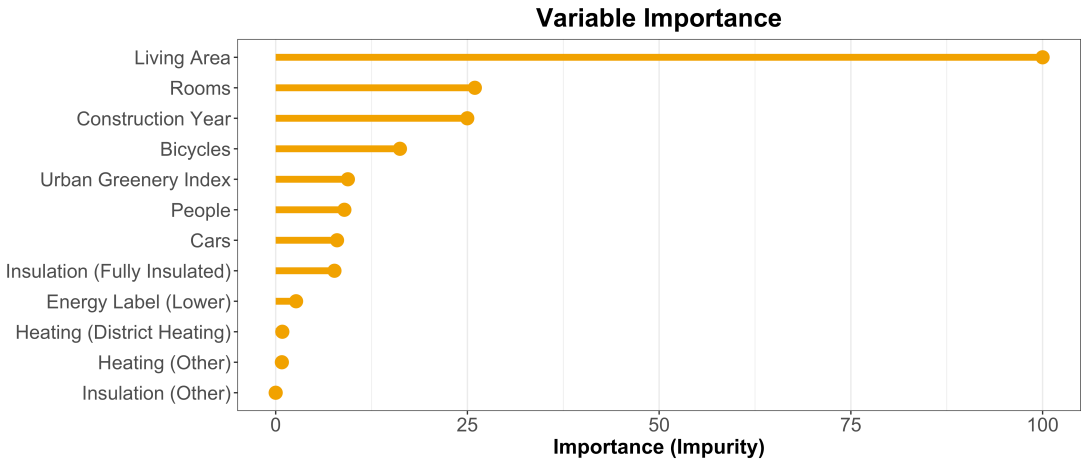

Figure 6.4: Variable Importance Plot for Random Forest model.

Strikingly, all four image-derived environmental features (Bicycles, UGI, People, and Cars) rank among the top seven most important variables. This high ranking underscores the substantial role these environmental factors play in explaining housing price variations, surpassing several traditional features.

The prominence of these environmental features, particularly the bicycle count and UGI, aligns with our earlier findings from the OLS models. This consistency across our modeling approaches reinforces the robustness of our results, providing evidence for the value of incorporating these image-derived features into housing price prediction models, suggesting that they capture important aspects of property valuation that might be missed by relying solely on traditional attributes.

SHAP and PDP Analysis

To investigate non-linear effects and potential thresholds, we employ SHAP plots and Partial Dependence Plots (PDPs) for each of our four main environmental features. Figure 6.5 presents a grid of these plots, allowing for direct comparison.

SHAP plots visualize how each feature impacts the model's predictions across its range of values. Each point represents a single prediction, with the x-axis showing the feature value and the y-axis indicating the SHAP value (impact on the prediction). The color of each point represents the count of people associated with that property, allowing us to observe potential interaction effects. Partial Dependence Plots (PDP), shown below each SHAP plot, illustrate the marginal effect of a feature on the predicted outcome, averaging over the effects of all other features.

We now analyze each environmental feature in detail:

Urban Greenery Index

The UGI analysis reveals a complex, non-linear relationship between greenery and housing prices. At very low UGI values, we observe slightly positive SHAP values, which particularly coincide with properties with high people counts associated with them. This counterintuitive finding likely reflects the high value of central, densely populated areas where even minimal greenery is rare. The model associates the presence of (almost) no

greenery with higher property values, not because of absence of the greenery itself, but due to the desirability of the location.

As UGI increases, we see a strong positive trend, especially in the mid-range values, indicating that increases in greenery generally correspond to higher property values. The partial dependence plot corroborates this, showing a sharp initial increase in predicted price as UGI rises from very low values. Notably, there's a clear threshold at around 40% UGI, after which the positive impact on price levels off, suggesting diminishing returns to increasing greenery beyond this point.

These findings diverge from those by Chen et al. (2020). While they found that homebuyers are only willing to pay more when UGI is sufficiently high, our results indicate the presence of a saturation point beyond which additional green space may not significantly increase property values. This nuanced relationship suggests that the value of urban greenery is highly context-dependent.

Presence of People

The people count exhibits a nuanced, non-linear relationship with housing prices. The SHAP plot shows a mix of positive and negative values at low people counts, transitioning to a general positive trend as the count increases. This trend is not uniform, with considerable dispersion in SHAP values at higher counts.

The partial dependence plot reveals an initial increase in predicted price as people count rises from very low values, followed by a plateau between approximately 2 and 4 people per image, and then a more gradual rise. This pattern suggests an initial premium for an "optimal" range of urban activity, and potentially a secondary premium for very busy, possibly commercial or central areas.

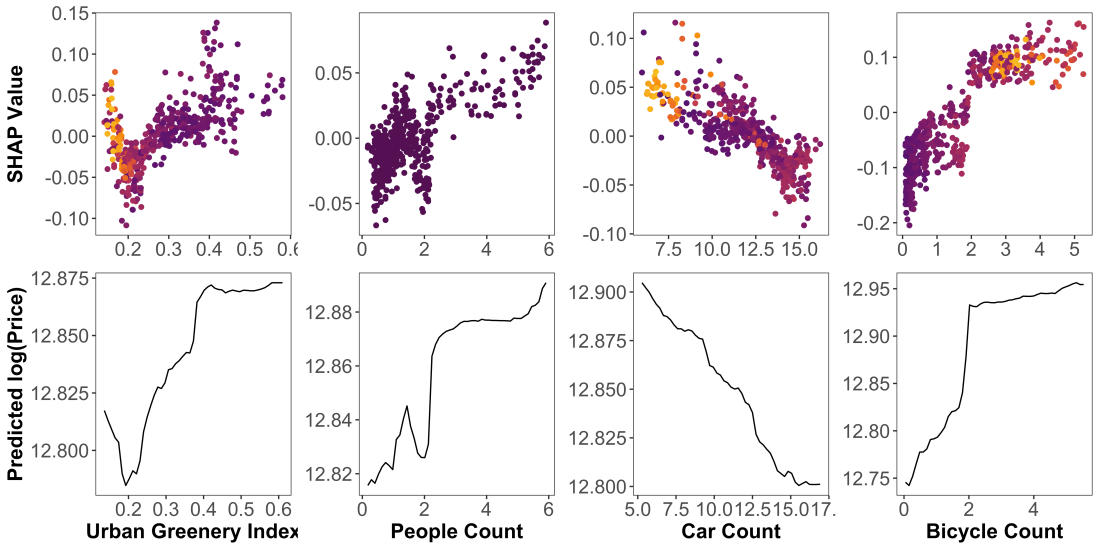

Figure 6.5: Grid of PDP and SHAP plots for all image features. The SHAP Plots are colored by the count of people associated with each property.

Number of Cars

Car count demonstrates a predominantly negative relationship with housing prices. The SHAP plot shows decreasing values as the number of cars increases, indicating a negative

impact on property values. This relationship appears relatively linear, with some dispersion at lower car counts. The partial dependence plot confirms this negative relationship, showing a consistent decrease in predicted price as car count increases.

While these findings support our initial hypothesis about the negative impact of cars on property values, we don't observe the hypothesized threshold effect. Notably, low vehicle densities appear to be inversely correlated with pedestrian presence, as evidenced by the lighter coloration of numerous high SHAP values. These data points, concentrated in the upper left quadrant of the plot, are indicative of high-value, centrally located properties in areas of high foot traffic. However, the negative effect of vehicle presence does not exhibit any plateau across the observed range, suggesting that the preference for lower car density is not limited to urban centers but appears to be consistent across the entire study area.

Number of Bicycles

Bicycle count exhibits a strong positive relationship with housing prices. The SHAP plot reveals a non-linear pattern, with a steep initial increase followed by a more gradual positive trend. Properties with high bicycle counts and high SHAP values often coincide with high people counts, indicating that areas with many bicycles and people are generally associated with higher property values.

The partial dependence plot confirms this relationship, showing a sharp initial increase in predicted price as the bicycle count rises from 0 to about 2 bicycles per image, followed by a much more gradual increase. This threshold effect suggests that the presence of bicycles is highly valued, possibly as an indicator of cycling infrastructure and neighborhood quality, but with diminishing marginal benefits beyond a certain point. These findings support our hypothesis about the positive impact of bicycles on property values while revealing a more nuanced relationship than initially expected.

Implications

As we move towards our conclusions, from our initial exploration of directional effects across housing types, through the investigation of spatial scales, to the uncovering of non-linear and threshold effects, we have demonstrated that the influence of urban environmental features on housing prices is far from straightforward. Our findings challenge simplistic assumptions about the value of environmental amenities, revealing that their impacts can be highly context-dependent and often non-linear. These findings collectively underscore the limitations of one-size-fits-all urban planning approaches and highlight the potential for targeted interventions to significantly impact property values and urban livability.

For instance, the strong positive effect of bicycle presence, with a clear threshold, suggests that even modest investments in cycling infrastructure could yield significant benefits. The varying impacts of greenery, human presence, and transportation-related factors across different property types and spatial scales highlight the need for differentiated approaches. The consistent outperformance of non-linear models, particularly Random Forests, in capturing these complex relationships underscores the limitations of traditional linear approaches in this domain.

7 Conclusions

This study set out to investigate how urban environmental factors, extracted from street-level imagery, influence residential property prices in Rotterdam. Our research contributes to the growing body of literature on real estate valuation by incorporating unstructured visual data to capture nuanced aspects of urban environments.

Our findings largely support our initial hypotheses while revealing some unexpected insights. The inclusion of image-derived environmental factors significantly enhanced the predictive accuracy of our housing price models, with the model specifications incorporating all features demonstrating a substantial improvement over the baseline models.

The directional effects of environmental features on housing prices varied, with some notable differences between houses and apartments. The Urban Greenery Index (UGI) and the presence of bicycles were identified as particularly influential factors that positively impact property values, with the former having a stronger effect on houses and the latter having a stronger effect on apartments.

Our analysis of spatial scales revealed that the impact of most environmental factors increases with distance, contradicting our hypothesis of decreasing influence. This suggests that broader neighborhood characteristics may be more influential than immediate surroundings in determining property values in Rotterdam. The non-linear and threshold effects uncovered in our analysis, particularly for the UGI and bicycle presence, support our fourth hypothesis and highlight the complexity of Rotterdam's urban housing market.

Methodologically, this study demonstrates the potential of leveraging street-level imagery and advanced computer vision techniques to enhance our understanding of urban property valuation. By extracting specific, interpretable environmental factors from images, we provide a more nuanced and comprehensive assessment of neighborhood characteristics.

These findings have important implications for urban planning and real estate practices. The strong positive effect of greenery on property values, particularly for houses, supports policies promoting the preservation and creation of urban green spaces. The negative impact of car presence and positive effect of bicycle presence on property values indicate urban living preferences that could inform future developments, encouraging designs that prioritize pedestrian and cyclist-friendly spaces over car-centric infrastructure. Furthermore, the increasing impact of environmental factors at larger spatial scales emphasizes the importance of cohesive, city-wide urban planning strategies.

This study demonstrates the significant potential of leveraging unstructured data and advanced analytical techniques to enhance our understanding of urban housing markets, potentially aiding informed decision making by city planners and policymakers. As Rotterdam continues to evolve, these findings provide a data-driven foundation for creating a more attractive and valuable urban environment.

8 Discussion

While this study has provided valuable insights, it is important to acknowledge its limitations and consider directions for future research.

Our reliance on listing prices rather than actual transaction data may introduce some bias. While listing prices generally reflect market values reasonably well, future studies could benefit from access to official transaction data from the Kadaster to validate and refine our findings. Given the dynamic nature of urban environments and evolving societal preferences, using data from multiple periods would be valuable for comparing how the influence of environmental factors on property values changes over time.

Our approach of using fixed radii for spatial association, while providing valuable insights, has limitations. The varying number of associated images at different radii may influence our results. Alternative methods, such as distance-weighted approaches, could provide a more nuanced understanding of spatial effects.

Future research could extend this study to other Dutch or international cities, providing valuable insights into the generalizability of our findings and distinguishing universal trends from location-specific effects.

Our study identifies correlations between environmental factors and property values but does not establish causation. Future research could employ causal inference techniques to better isolate these effects. Studying the impact of new green space developments or cycling infrastructure projects on nearby property values over time could provide stronger evidence of causal relationships.

Comparing our method of extracting specific environmental factors with end-to-end frameworks that use images to represent overall neighborhood attractiveness could be valuable. While end-to-end approaches might be less interpretable, such comparisons could shed light on the trade-offs between interpretability and predictive power in property valuation models.

Furthermore, future studies could use our methodology to assess the impact of specific urban interventions or policy changes. For example, researchers could investigate how the introduction of a new park or a major change in transportation infrastructure affects property values in surrounding areas, providing valuable feedback for urban planning decisions.

While our study showcases the potential of leveraging street-level imagery and advanced computer vision techniques to enhance our understanding of urban property valuation, it also opens up a wealth of new questions and research directions. As cities continue to evolve and face new challenges, from climate change to changing demographics, the need for data-driven insights to inform urban planning and real estate decisions will only grow. We hope that the scalable and objective methods developed in this study serve as a foundation for further research and provide valuable insights for urban planners, policymakers, and real estate professionals.

Acknowledgments

I would like to express my sincere gratitude to my thesis supervisor, Professor Rommert Dekker, for his invaluable guidance, consistent support, and insightful feedback throughout this research process. His engagement and expertise were instrumental in shaping this study.

I am deeply indebted to Francisco Garrido-Valenzuela from TU Delft for his generous assistance in providing access to the extensive Google Street View image dataset. His willingness to share resources and offer insights into working with this data significantly enhanced the scope and was instrumental in achieving the comprehensive approach of this research.

This thesis would not have been possible without their contributions.

Supplementary Figures

	Darknet-53 Architecture			
	Type	Filters	Size	Output
	Convolutional	32	3 x 3	256 x 256
	Convolutional	64	3 x 3 / 2	128 x 128
	Convolutional	32	1 x 1	
1x	Convolutional	64	3 x 3	
	Residual			128 x 128
	Convolutional	128	3 x 3 / 2	64 x 64
	Convolutional	64	1 x 1	
2x	Convolutional	128	3 x 3	
	Residual			64 x 64
	Convolutional	256	3 x 3 / 2	32 x 32
	Convolutional	128	1 x 1	
8x	Convolutional	256	3 x 3	
	Residual			32 x 32
	Convolutional	512	3 x 3 / 2	16 x 16
	Convolutional	256	1 x 1	
8x	Convolutional	512	3 x 3	
	Residual			16 x 16
	Convolutional	1024	3 x 3 / 2	8 x 8
	Convolutional	512	1 x 1	
4x	Convolutional	1024	3 x 3	
	Residual			8 x 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Supplementary Figure S8.1: Darknet-53 Network Design, which is the YOLO-v3 backbone.

Variable	Coefficient	95% CI	**Adjusted GVIF**
(Intercept)	10.657***	(10.247, 11.068)	
Living Area (m²)	0.009***	(0.008, 0.009)	1.7
Energy Label			1.2
Higher (A+, A, B, C)	—	—	
Lower (D, E, F, G)	-0.050***	(-0.062, -0.038)	
Rooms	-0.012***	(-0.018, -0.005)	1.7
Construction Year	0.001***	(0.000, 0.001)	1.5
Insulation			1.1
Double Glass	—	—	
Fully Insulated	0.171***	(0.155, 0.187)	
Other	0.007	(-0.006, 0.021)	
Heating			1.2
Central Heating	—	—	
District Heating	0.044***	(0.027, 0.060)	
Other	-0.047***	(-0.062, -0.032)	
UGI	0.490***	(0.410, 0.571)	1.5
People	0.000	(-0.007, 0.008)	2.0
Cars	-0.021***	(-0.024, -0.018)	1.4
Bicycles	0.091***	(0.084, 0.098)	1.8
R^2	0.777		
R^2 *(adj.)*	0.776		
p-value	<0.001		
N	6649		

[1] *p<0.05;* ***p<0.01;*** p<0.001

[2] CI = Confidence Interval, GVIF = Generalized Variance Inflation Factor

[3] GVIF^[1/(2*df)]

Supplementary Figure S8.2: The table presents the adjusted generalized variance inflation factor (aGVIF) for each variable in the OLS model. Variables with high aGVIF values are typically considered influential and may pose potential issues for the model's performance. However, none of the variables in our model exhibit such problematic levels of multicollinearity.

Supplementary Figure S8.3: The impact of construction year on house prices is U-shaped, where the age of both very old and very new houses is being valued positively.

References

Abidoye, R. B. and Chan, A. P. 2018. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network, *Pacific Rim Property Research Journal*, vol. 24, no. 1, 71–83

An, S., Jang, H., Kim, H., Song, Y., and Ahn, K. 2023. Assessment of street-level greenness and its association with housing prices in a metropolitan area, *Scientific reports*, vol. 13, no. 1, 22577

Badrinarayanan, V., Kendall, A., and Cipolla, R. 2016. SegNet: A deep convolutional encoder-decoder architecture for image segmentation

Bency, A. J., Rallapalli, S., Ganti, R. K., Srivatsa, M., and Manjunath, B. 2017. Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery, pp. 320–29, in *2017 IEEE winter conference on applications of computer vision (WACV)*, IEEE

Bond, M., Seiler, M., Seiler, V., and Blake, B. 2000. Uses of websites for effective real estate marketing, *Journal of Real Estate Portfolio Management*, vol. 6, no. 2, 203–11

Bourassa, S. C., Peng, V. S., et al. 1999. Hedonic prices and house numbers: The influence of feng shui, *International Real Estate Review*, vol. 2, no. 1, 79–93

Boyle, M. and Kiel, K. 2001. A survey of house price hedonic studies of the impact of environmental externalities, *Journal of real estate literature*, vol. 9, no. 2, 117–44

Breiman, L. 2001. Random forests, *Machine learning*, vol. 45, 5–32

Butler, R. V. 1982. The specification of hedonic indexes for urban housing, *Land economics*, vol. 58, no. 1, 96–108

Chau, K. W. and Chin, T. 2003. A critical review of literature on the hedonic price model, *International Journal for Housing Science and its applications*, vol. 27, no. 2, 145–65

Chau, K., Ma, V., and Ho, D. 2001. The pricing of 'luckiness' in the apartment market, *Journal of Real Estate Literature*, vol. 9, no. 1, 29–40

Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., and Girdhar, R. 2022. Masked-attention mask transformer for universal image segmentation, pp. 1290–99, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*

Chen, L., Yao, X., Liu, Y., Zhu, Y., Chen, W., Zhao, X., and Chi, T. 2020. Measuring

impacts of urban environmental elements on housing prices based on multisource data—a case study of shanghai, china, *ISPRS Int. J. Geo-Inf.*, vol. 9, 106

Clark, D. E. and Herrin, W. E. 2000. The impact of public school attributes on home sale prices in california, *Growth and change*, vol. 31, no. 3, 385–407

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. 2016. The cityscapes dataset for semantic urban scene understanding, pp. 3213–23, in *Proceedings of the IEEE conference on computer vision and pattern recognition*

Doan, Q. C., Chen, C., He, S., and Zhang, X. 2024. How urban air quality affects land values: Exploring non-linear and threshold mechanism using explainable artificial intelligence, *Journal of Cleaner Production*, vol. 434, 140340

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv preprint arXiv:2010.11929*, Advance Access published 2020

Espey, M. and Lopez, H. 2000. The impact of airport noise and proximity on residential property values, *Growth and change*, vol. 31, no. 3, 408–19

Fletcher, M., Gallimore, P., and Mangan, J. 2000. The modelling of housing submarkets, *Journal of Property Investment & Finance*, vol. 18, no. 4, 473–87

Forrest, D., Glen, J., and Ward, R. 1996. The impact of a light rail system on the structure of house prices: A hedonic longitudinal study, *Journal of Transport economics and Policy*, 15–29

Fu, X., Jia, T., Zhang, X., Li, S., and Zhang, Y. 2019. Do street-level scene perceptions affect housing prices in chinese megacities? An analysis using open access datasets and deep learning, *PloS one*, vol. 14, no. 5, e0217505

Garrido-Valenzuela, F., Cats, O., and Cranenburgh, S. van. 2023. Where are the people? Counting people in millions of street-level images to explore associations between people's urban density and urban characteristics, *Computers Environment and Urban Systems*, vol. 102, 101971

Garrod, G. D. and Willis, K. G. 1992. Valuing goods' characteristics: An application of the hedonic price method to environmental attributes, *Journal of Environmental management*, vol. 34, no. 1, 59–76

Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Aiden, E. L., and Fei-Fei, L. 2017. Using deep learning and google street view to estimate the demographic makeup of neighborhoods across the united states, *Proceedings of the National Academy of Sciences*, vol. 114, no. 50, 13108–13

Goodman, A. C. 1978. Hedonic prices, price indices and housing markets, *Journal of urban economics*, vol. 5, no. 4, 471–84

Guite, H. F., Clark, C., and Ackrill, G. 2006. The impact of the physical and urban environment on mental well-being, *Public health*, vol. 120, no. 12, 1117–26

Hamilton, J. and Gunesh, R. 2003. Incorporating customer interface-marketing design elements to leverage strategic positioning in the on-line real estate industry, in *Hawaii international conference on business*, Citeseer

Harrison Jr, D. and Rubinfeld, D. L. 1978. Hedonic housing prices and the demand for clean air, *Journal of environmental economics and management*, vol. 5, no. 1, 81–102

Hong, J., Choi, H., and Kim, W. 2020. A house price valuation based on the random forest approach: The mass appraisal of residential property in south korea, *International Journal of Strategic Property Management*, vol. 24, no. 3, 140–52

Jim, C. Y. and Chen, W. Y. 2006. Impacts of urban environmental elements on residential housing prices in guangzhou (china), *Landscape and Urban Planning*, vol. 78, no. 4, 422–34

Kolbe, J. and Wüstemann, H. 2014. Estimating the value of urban green space: A hedonic pricing analysis of the housing market in cologne, germany: SFB 649 Discussion Paper

Kostic, Z. and Jevremovic, A. 2020. What image features boost housing market predictions?, *IEEE Transactions on Multimedia*, vol. 22, no. 7, 1904–16

Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, vol. 25

Lancaster, K. J. 1966. A new approach to consumer theory, *Journal of political economy*, vol. 74, no. 2, 132–57

Law, S., Paige, B., and Russell, C. 2019. Take a look around: Using street view and satellite images to estimate house prices, *arXiv*, Advance Access published 2019

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, 2278–2324

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. 2014. Microsoft coco: Common objects in context, pp. 740–55, in *Computer vision–ECCV 2014: 13th european conference, zurich, switzerland, september 6-12, 2014, proceedings, part v 13*, Springer

Linneman, P. 1980. Some empirical results on the nature of the hedonic price function for the urban housing market, *Journal of urban economics*, vol. 8, no. 1, 47–68

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows, pp. 10012–22, in *Proceedings of the IEEE/CVF international conference on computer vision*

Lundberg, S. M. and Lee, S.-I. 2017. A unified approach to interpreting model predictions, *Advances in neural information processing systems*, vol. 30

Malpezzi, S. et al. 2003. Hedonic pricing models: A selective and applied review, *Housing economics and public policy*, vol. 1, 67–89

Mason, C. and Quigley, J. M. 1996. Non-parametric hedonic housing prices, *Housing studies*, vol. 11, no. 3, 373–85

McMillan, D., Jarmin, R., and Thorsnes, P. 1992. Selection bias and land development in the monocentric model, *Journal of Urban Economics*, vol. 31, 273–84

Michaels, R. G. and Smith, V. K. 1990. Market segmentation and valuing amenities with hedonic models: The case of hazardous waste sites, *Journal of Urban Economics*, vol. 28, no. 2, 223–42

Mok, H. M., Chan, P. P., and Cho, Y.-S. 1995. A hedonic price model for private properties in hong kong, *The Journal of Real Estate Finance and Economics*, vol. 10, 37–48

Muhr, V., Despotovic, M., Koch, D., Döller, M., and Zeppelzauer, M. 2017. Towards automated real estate assessment from satellite images with CNNs., pp. 14–23, in *Forum media technology*

Mu, J., Wu, F., and Zhang, A. 2014. Housing value forecasting based on machine learning methods, p. 648047, in *Abstract and applied analysis*, Wiley Online Library

Naik, U. P., Rajesh, V., Kumar, R., et al. 2021. Implementation of YOLOv4 algorithm for multiple object detection in image and video dataset using deep learning and artificial intelligence for urban traffic video surveillance application, pp. 1–6, in *2021 fourth international conference on electrical, computer and communication technologies (ICECCT)*, IEEE

Niels, G. 2018. Funda-mentals of article 102: A dominant platform, but no abuse, *Oxera Agenda*, Advance Access published 2018

Palmquist, R. B. 1989. Valuing localized externalities, *Revealed Preference Approaches to Environmental Valuation Volumes I and II*, Advance Access published 1989

Petermann, J. 2021. DEVELOPMENT OF REAL ESTATE MARKETING–TRENDS FOR THE FUTURE., *Marketing Science & Inspirations*, vol. 16, no. 4

Potrawa, T. and Tetereva, A. 2022. How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market, *Journal of Business Research*, vol. 144, 50–65

Poursaeed, O., Matera, T., and Belongie, S. 2018. Vision-based real estate price estimation, *arXiv*, Advance Access published 2018

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. 2016. You only look once: Unified,

real-time object detection, pp. 779–88, in *Proceedings of the IEEE conference on computer vision and pattern recognition*

Redmon, J. and Farhadi, A. 2018. Yolov3: An incremental improvement, *arXiv preprint arXiv:1804.02767*, Advance Access published 2018

Richardson, H. W., Vipond, J., and Furbey, R. A. 1974. Determinants of urban house prices, *Urban studies*, vol. 11, no. 2, 189–99

Rodriguez, M. and Sirmans, C. 1994. Quantifying the value of a view in single-family housing markets, *Appraisal Journal*, vol. 62, 600–600

Rosen, S. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition, *Journal of political economy*, vol. 82, no. 1, 34–55

Semnani, S. J. and Rezaei, H. 2021. House price prediction using satellite imagery, *arXiv preprint arXiv:2105.06060*, Advance Access published 2021

Shaw, J. 2020. Platform real estate: Theory and practice of new urban real estate markets, *Urban Geography*, vol. 41, no. 8, 1037–64

Sihi, D. 2018. Home sweet virtual home: The use of virtual and augmented reality technologies in high involvement purchase decisions, *Journal of Research in Interactive Marketing*, vol. 12, no. 4, 398–417

Simonyan, K. and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*, Advance Access published 2014

Sirmans, S., Macpherson, D., and Zietz, E. 2005. The composition of hedonic pricing models, *Journal of real estate literature*, vol. 13, no. 1, 1–44

So, H. M., Tse, R. Y., and Ganesan, S. 1997. Estimating the influence of transport on house prices: Evidence from hong kong, *Journal of property valuation and investment*, vol. 15, no. 1, 40–47

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. 2014. Going deeper with convolutions

Tyrväinen, L. 1997. The amenity value of the urban forest: An application of the hedonic pricing method, *Landscape and Urban planning*, vol. 37, nos. 3-4, 211–22

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

Xu, X., Qiu, W., Li, W., Liu, X., Zhang, Z., Li, X., and Luo, D. 2022. Associations between street-view perceptions and housing prices: Subjective vs. Objective measures using computer vision and machine learning techniques, *Remote Sensing*, vol. 14, no. 4, 891

Yazdani, M. and Raissi, M. 2023. Real estate property valuation using self-supervised vision transformers, *arXiv preprint arXiv:2302.00117*, Advance Access published 2023

Yencha, C. 2019. Valuing walkability: New evidence from computer vision methods, *Transportation research part A: policy and practice*, vol. 130, 689–709

Ye, Y., Xie, H., Fang, J., Jiang, H., and Wang, D. 2019. Daily accessed street greenery and housing price: Measuring economic performance of human-scale streetscapes via new urban data, *Sustainability*, vol. 11, 1741

You, Q., Pang, R., Cao, L., and Luo, J. 2017. Image-based appraisal of real estate properties, *IEEE transactions on multimedia*, vol. 19, no. 12, 2751–59

Zhang, Y. and Dong, R. 2018. Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in beijing, *ISPRS Int. J. Geo-Inf.*, vol. 7, 104

Zhao, C., Ogawa, Y., Chen, S., Oki, T., and Sekimoto, Y. 2023. Quantitative land price analysis via computer vision from street view images, *Engineering Applications of Artificial Intelligence*, vol. 123, 106294