Cities 174 (2026) 107066

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities

A utility-based spatial analysis of residential street-level conditions a case study of Rotterdam

Sander van Cranenburgh [a,*], Francisco Garrido-Valenzuela [a]

[a] CityAI lab, Transport and Logistics Group, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Residential location choice
Urban environment
Discrete choice models
Segmentation
Street-level images

ABSTRACT

This study sheds light on how utility derived from street-level conditions is spatially distributed, from a residential location choice perspective, at a city-wide scale. Unlike previous studies that analyse perceptions of urban environments from street-level imagery, this work maps preferences—that is, the utility residents derive from observable street-level conditions. To this end, we first develop a residential location discrete choice model that builds on two premises: (1) street-level images effectively capture street-level conditions, and (2) state-of-the-art segmentation models can extract salient information from these images and convert them into structured (i.e. tabular) data. We then apply the model to over 200 thousand geo-tagged street-level images of Rotterdam (the Netherlands) to map how utility derived from street-level conditions varies across the city. Results show strong local variation, with conditions changing rapidly even within neighbourhoods, and reveal that high real-estate prices in the city centre cannot primarily be attributed to attractive street-level conditions. As a secondary methodological contribution, the paper integrates foundation segmentation models into discrete choice analysis. Unlike conventional segmentation approaches limited to predefined object classes, our pipeline leverages prompt-based detection (GroundingDINO + SAM) to identify novel and more granular categories (e.g. transformer houses, shrubs vs. trees) overlooked in standard datasets. This integration enables a richer, fine-grained quantification of street-level conditions and demonstrates how visual information can be systematically embedded into residential location choice models. As such, this paper's findings and methodological contribution pave the way for further studies to explore integrating street-level conditions in urban planning.

1. Introduction

Residential location choices shape the infrastructure and functionality of cities. Specifically, the residential location choices that individuals and households make have significant implications for transportation systems, housing markets, and urban planning (Cox & Hurtubia, 2021). Therefore, formulating policies to address urban challenges such as sprawl, housing affordability, and spatial inequality requires a thorough understanding of the factors influencing residential location choices (Pagliara & Wilson, 2010).

Residential location choices are commonly analysed using so-called Discrete Choice Models (DCMs) (Bostanara, Siripanich, & Rashidi, 2024; Guevara & Ben-Akiva, 2006; Hunt, 2010; McFadden, 1977; Pérez, Martínez, & Ortúzar, 2003). These disaggregate models are grounded in random utility theory, which posits that individuals select the alternative that maximises their utility from a finite set of discrete alternatives. Each alternative is conceptualised as a bundle of attributes, such as cost, travel time, or quality (Ben-Akiva & Lerman, 1985; Lancaster, 1966). Researchers estimate the parameters of the utility functions of DCMs by confronting them with empirical data that involves trade-offs across the attributes. These parameters quantify the relative importance of each attribute in the decision-making. Three main types of factors are found to be of importance to residential location choice behaviour (Schirmer, Van Eggermond, & Axhausen, 2014), namely (1) *Travel and accessibility-related factors*, such as the commute mode, commuting time, and distances to amenities like schools, stores, hospitals and playgrounds (e.g. Frenkel, Bendit, & Kaplan, 2013); (2) *Socioeconomic environments*, such as income levels, ethnic distribution, age and education level (e.g. Clark & Ledwith, 2007); and, (3) *Built environments characteristics and street-level conditions*, which together describe the physical and visual qualities of a residential area. Built environment characteristics typically include features such as building density (Waddell, 2006; Xue & Yao, 2022), housing typology, and street layout (Pinjari, Pendyala, Bhat, & Waddell, 2007; Yang, Ding, Ju, & Yu, 2021). Street-level conditions

* Corresponding author at: Transport and Logistics group, Faculty of Technology Policy and Management, Jaffalaan 5, 2286 BX, Delft, the Netherlands.
E-mail address: s.vancranenburgh@tudelft.nl (S. van Cranenburgh).

https://doi.org/10.1016/j.cities.2026.107066
Received 23 July 2025; Received in revised form 22 October 2025; Accepted 30 March 2026
Available online 13 April 2026

refer to perceptual and interpretative assessments of the environment, such as the overall sense of safety, cleanliness, or child-friendliness (Gong, van den Berg, Dane, & Arentze, 2025). These conditions emerge from the visual composition of the street scene as a whole and are typically shaped by interpreting the combination of visual components.

Notwithstanding the acknowledged importance of street-level conditions to residential location choices (e.g. Giles-Corti et al., 2013), they are often incorporated in a simplistic manner –or overlooked entirely– in many residential location choice studies (Schirmer et al., 2014). This oversight largely stems from data practices. Residential location choice models have traditionally been built using census data, which do not contain detailed information about the built environment and street-level conditions. Unlike traditional tabular census and cadastre data, street-level images are particularly adept at encoding information about street-level conditions. The widespread use of street-level images on housing platforms and real-estate agency websites attests to the power of images to describe and convey information about street-level conditions as well as their importance to residential location choices (Seiler, Madhavan, & Liechty, 2012). Importantly, nowadays, images containing information on street-level conditions are widely available from tech firms like Google, Apple, and Baidu's map services. In addition, recent advances in computer vision enable the efficient extraction of information encoded in images. For instance, segmentation models can be used to identify and quantify components such as trees, roads, and sky, providing structured data from unstructured imagery. Since these images and the tools to extract information from them became widely accessible, researchers have extensively utilised them to analyse and understand urban environments (Biljecki & Ito, 2021; Garrido-Valenzuela, Cats, & van Cranenburgh, 2023; Zhang et al., 2024).

However, street-level images have not yet been used to examine how utility derived from street-level conditions is spatially distributed from a residential location choice perspective. As a result, little is known about how the utility associated with these conditions varies across neighbourhoods within cities. This knowledge gap hinders urban planners from making informed context-specific decisions to enhance the attractiveness of local areas from a residential perspective. Effective policies to improve the street-level conditions may vary from place to place. For instance, a highly urbanised, densely populated area may require more open views, while a suburb may especially benefit from fewer parked cars.

This study aims to fill this gap. It aims to shed light on how utility derived from street-level conditions is spatially distributed, from a residential location choice perspective, at a city-wide scale. Additionally, it aims to examine the factors shaping this distribution. To achieve our research objectives, we first develop a residential location discrete choice model that explicitly incorporates street-level conditions and then apply the developed model to a large urban area: the city of Rotterdam. Rotterdam, the second largest city in the Netherlands, with about 670 thousand inhabitants, boasts a diverse array of neighbourhoods (Custers & Willems, 2024; Doucet & Koenders, 2018), making it an ideal case study. More specifically, we estimate our residential location choice model based on data from a recent residential stated choice experiment, in which respondents faced trade-offs between street-level conditions –presented through street-level images– alongside two numeric attributes: monthly housing cost and commute travel time. As discrete choice models require structured data, we first converted the information encoded in the images about the street-level conditions into structured (i.e. tabular) data using state-of-the-art segmentation models. We then apply the estimated model to predict the spatial distribution of utility derived from street-level conditions across the entire city of Rotterdam. This is done by calculating the utility associated with street-level conditions captured in over 200k geo-tagged street-level images of Rotterdam. Then, we aggregate the results at the postal code level to generate a map that visualises the spatial distribution of utility derived from street-level conditions. Finally, we examine how the mapped utility distribution corresponds to existing indicators of residential quality, such as real-estate prices, liveability scores, and beauty perception.

The primary contribution of this paper is substantive: it presents insights into residential location *preferences* of urban environments. Thereby, it complements previous research on *perceptions* of the urban environment based on street-level imagery (e.g. Gu, Xu, Gong, & Liu, 2025; Rui, 2023; Zhang et al., 2018). As highlighted in Van Cranenburgh and Garrido-Valenzuela (2025), perceptions and preferences are closely related but distinct concepts. Preferences are grounded in the theory of choice behaviour (Lancaster, 1966; Luce, 1959; Samuelson, 1948) and govern what people choose (typically leading to economic demand). Perceptions are people's subjective interpretations of sensory stimuli, which can, but do not necessarily, influence individuals' choices. Our work also differs from studies using street-level imagery to predict housing prices (e.g. Law, Paige, & Russell, 2019). Whereas house-price models estimate market valuations, which are driven by housing attributes (e.g., dwelling size, interior quality, or design), public-space attributes (e.g., greenery, cleanliness, accessibility) and broader market expectations, our utility-based approach *isolates* the value residents derive specifically from street-level conditions. Consequently, our study provides explanatory insights that are directly actionable for urban planners and local governments seeking to enhance the quality and attractiveness of residential environments.

As a secondary methodological contribution, this paper pioneers the use of foundation segmentation models for urban analysis. Unlike conventional segmentation models, which are trained on datasets with predefined classes (e.g., car, road, wall), foundation segmentation models enable image segmentation guided directly by text prompts. This capability allows researchers to investigate categories beyond those included in standard segmentation datasets such as Cityscapes and ADE20K (Cordts et al., 2016; Zhou et al., 2017). As an illustration of this enhanced flexibility, we included prompts targeting electrical transformer houses. These structures offer a compelling case for three key reasons. First, despite their frequent presence in urban environments, transformer houses are typically overlooked by traditional segmentation models due to the absence of relevant labels in standard datasets. Second, the ongoing energy transition is driving the need for a substantial increase in the number of transformer houses to support grid expansion. Third, the siting of new transformer houses often sparks opposition from residents, indicating that their presence in the street-level environment may generate disutility.

The remaining part of this paper is organised as follows. Section 2 describes the method. It presents the model development and model application pipelines. As part of the model development, we compare the segmentations obtained using the conventional and foundation-based segmentation models in light of our objective to develop a residential location choice behaviour. Section 3 presents the results. Specifically, Section 3.1 presents the model development results, and Section 3.2 presents the main substantive findings. Finally, the paper ends with a conclusion and a discussion of the limitations and avenues for future research (Section 4).

2. Method

Fig. 1 shows the overview of our method. It comprises two main parts. Part 1 involves two steps (A and B) and develops the residential location choice model, and Part 2 involves four steps (A to D) and applies the developed residential location choice model to the city of Rotterdam. Next, we discuss each part in more detail.

2.1. Development of the residential location choice model

2.1.1. Residential location choice data

Part 1 of our method starts with residential location choice data, which are needed to develop the residential location choice model. For

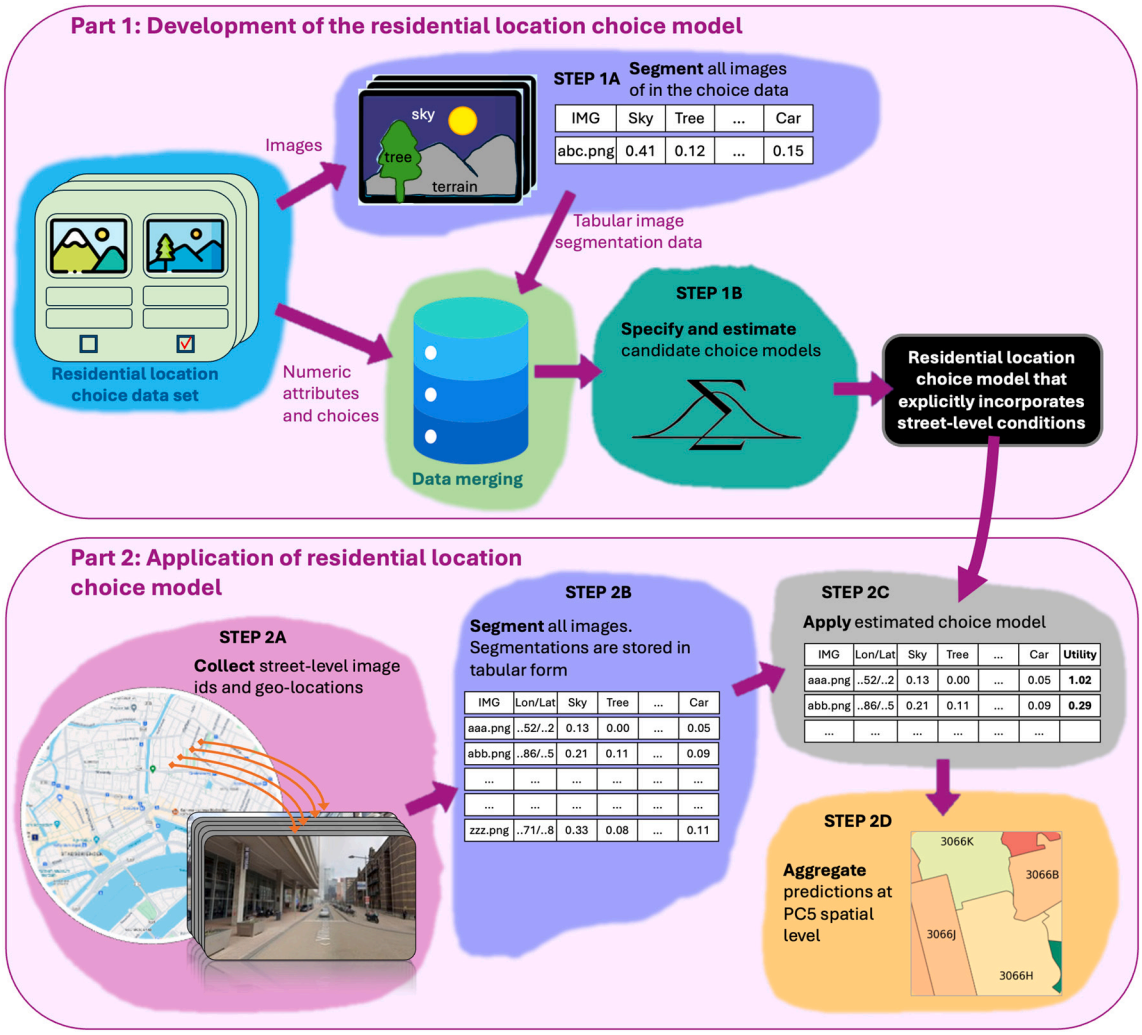


Fig. 1. Overview of the method.

this purpose, we use the stated choice data collected by Van Cranenburgh and Garrido-Valenzuela (2025) which have been made openly available.[1] In their experiment, respondents had to make trade-offs between street-level conditions, presented using an image, and two numeric attributes: monthly housing cost and commute travel time. Fig. 2 shows a screenshot of a choice task from their stated choice experiment. Van Cranenburgh and Garrido-Valenzuela (2025) drew the images randomly from a large database of street-level images. The experiment was conducted by a largely representative sample ($N = 800$) of the Dutch 18+ year-old population. Each participant was presented with 15 choice tasks. After cleaning, the data set contains 11,731 choice observations in which 7,341 unique street-level images are used. Furthermore, the data set has been split into a training (henceforth referred to as the estimation set) and a test set (henceforth referred to as the hold-out set) in such a way that the images do not overlap between the two sets.

2.1.2. Step 1A: image segmentation

The first step involves segmenting the images used in the stated choice experiment. The goal of this segmentation process is to extract explicit components from the images that are believed to characterise street-level conditions. In doing so, we adopt the typology proposed by Garrido-Valenzuela (2026), which distinguishes between *components* and *conditions*. Explicit components refer to tangible elements of the visual scene, including physical objects (e.g., cars, people, trees), structures (i.e., recognisable patterns and relationships between objects), and spatial concepts (e.g., sky, horizon, boundaries). Conditions, by contrast, refer to interpretative assessments of the environment. These segmented components are subsequently converted into structured data suitable for estimating discrete choice models. To achieve this, we employ and compare two state-of-the-art segmentation models.

The first is a task-specific segmentation model, developed by Meta AI, called the Mask2Former model (Cheng et al., 2022), which we apply using pre-trained weights based on the Cityscapes dataset. This model is well-suited for urban scenes as it is trained to recognise 19 classes, including buildings, roads, vegetation, cars, and traffic lights (Cordts et al., 2016).

The second segmentation model is a foundation segmentation model, which we use to demonstrate how recent language-vision models can flexibly extract visual information from images beyond predefined classes. Unlike conventional, task-specific segmentation models, such as Mask2former –which are trained on datasets with a predefined set of classes – foundation segmentation models enable image segmentation guided directly by open-ended natural-language prompts. The main advantage of this capability is that it allows the investigation of classes beyond those included in standard segmentation datasets such as Cityscapes. To illustrate this flexibility, we included prompts targeting transformer houses; an object type that plays an increasing role in the urban landscape but is typically omitted from standard segmentation

[1] https://github.com/TUD-CityAI-Lab/Utility-based-street-level-conditions

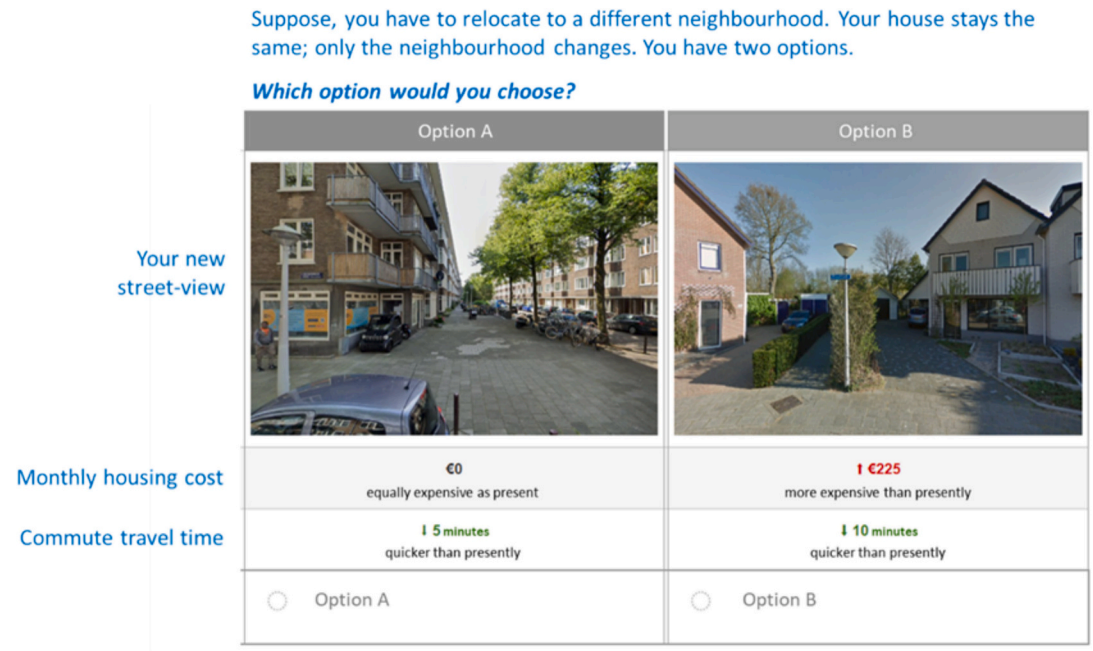


Fig. 2. Screenshot of the residential stated choice experiment. Adopted from Van Cranenburgh and Garrido-Valenzuela (2025).

taxonomies.

Specifically, our foundation segmentation pipeline combines two models: an object detection model called GroundingDINO and a segmentation model called Segment Anything Model (SAM). GroundingDINO (Liu et al., 2023) is a zero-shot object detection model that unifies transformer-based detection (Zhang et al., 2022) with grounded language–vision pre-training. In practice, this means that rather than manually assigning labels to images or using a predefined set of detection classes, we provide text prompts formulated in natural language, such as, "car", "tree", "waste bin", or "transformer house". The model then predicts bounding boxes for image regions that semantically match these textual descriptions, drawing on its pre-trained multimodal embeddings. For each prompt, GroundingDINO outputs a ranked list of detections with confidence scores between 0 and 1. We adopted a box threshold of 0.30 and a text threshold of 0.25, consistent with the recommended range in the official GroundingDINO implementation. These values provided a good balance between missed detections and spurious predictions after pilot testing on a representative image subset. Each detection's bounding box is then used to initialise the Segment Anything Model (SAM) (Kirillov et al., 2023), which performs pixel-accurate segmentation inside the bounding box. SAM delineates the detected object and returns the number of pixels, which we use to derive component shares (e.g. fraction of image covered by vegetation) and object counts (e.g. number of cars). Because both GroundingDINO and SAM are foundation models trained on hundreds of millions of image–text pairs, they operate without any additional manual labelling or fine-tuning. No task-specific classification layer is required: all categories are defined purely through the text prompts. This makes the method flexible and scalable, allowing the extraction of classes absent from conventional segmentation datasets – such as transformer houses or waste bins – simply by providing new prompts.

We applied both segmentation pipelines to all images used in the stated choice experiment. We extracted the proportion of pixels assigned to each class from the resulting segmentation. Additionally, we recorded object counts for certain countable components, such as cars or bicycles. The predefined set of classes and the set of prompts used in the foundation segmentation model are listed in Table 1. In designing the prompts, we aligned with the relevant Cityscapes classes while also refining certain ones that are handled in a relatively coarse manner.

Table 1
Segmentation classes and prompts.

Class/ Prompt	Segmentation model pretrained on Cityscapes	Foundation segmentation model	Treated as countable
Bicycle	X	X	
Building	X	X	
Bus	X	X	
Car	X	X	X
Fence	X	X	
Grass		X	
House		X	
Motorcycle	X	X	X
Person	X	X	X
Plants		X	
Pole	X	X	
Rider	X		
Road	X	X	
Shrubs		X	
Sidewalk	X	X	
Sky	X	X	
Terrain	X	X	
Traffic light	X	X	X
Traffic sign	X	X	X
Train	X		
Transformer house		X	X
Trees		X	
Truck	X	X	X
Vegetation	X		
Wall	X		
Waste bin		X	X
Water		X	
Total #	**19**	**23**	**8**

Thereby, we aim to capture more nuanced features of the street-level environment that may be important in shaping residential location choices. In particular, we used the classes 'Grass', 'Plants', 'Shrubs', and 'Trees' to refine the segmentation of the Cityscapes class 'Vegetation'. Additionally, we included the class 'Water', based on the expectation that residents may derive positive utility from the presence of a pond or canal in their street view. Furthermore, the two Cityscapes classes, 'Rider' and 'Train', were a priori discarded based on the expectation that they are not relevant in shaping location choices or are rarely present in imagery. Finally, as explained above, we added 'Transformer house' to

illustrate the enhanced capability of foundation segmentation models. To improve detection accuracy, we used two prompts: 'Transformer house' and 'Electric substation'. We also included the prompt 'Waste bin', as initial trials revealed that omitting this class led to frequent misclassification of waste bins as transformer houses.

While a comprehensive comparison between the segmentation approaches lies beyond the scope of this paper, we find strong correspondence between the two models for the majority of classes, particularly for dominant urban elements such as buildings, roads, cars, and sky. Since the accuracy of the Mask2Former model, trained on the Cityscapes dataset, is well-established in the literature (Cheng, Misra, Schwing, Kirillov, & Girdhar, 2022), this close alignment provides additional confidence in the validity of the GroundingDINO + SAM segmentations. Fig. 3 illustrates this further using three images. First, there is substantial overlap between the outputs, indicating that both models are broadly aligned in identifying the major components of the street-level environment. Second, the foundation model captures more refined classes –for example, distinguishing between grass, plants, and trees– and is capable of detecting transformer houses, which the Mask2Former model fails to identify. However, it mislabels the left-hand side transformer house as a waste bin, likely due to their subtle visual similarities. This pattern is confirmed through additional inspection of images containing transformer houses.

2.1.3. Step 1B: discrete choice model specification and estimation

We employ a Random Utility Maximisation (RUM) Multinomial Logit (MNL) model rather than more advanced mixture models, such as Mixed Logit or Latent Class models. While these advanced models often achieve superior fit in estimation, their advantage does not carry over to the prediction of future choices of "new" individuals (see Hess & van Cranenburgh, 2025). The MNL model, therefore, offers an equally accurate, yet more parsimonious and computationally efficient specification for our application. Eq. 1 presents the model, where U_{in} denotes the total utility that decision maker n derives from residential alternative i. This utility consists of two components: an observed part V_{in}, and an unobserved part ε_{in}, which captures idiosyncratic influences not included in the model. The observed utility V_{in} is specified as a linear-additive function of attributes m, which include both the street-level components extracted from the segmented images (counts and shares) and the numeric attributes housing cost and commute travel time. Under the assumption that ε_{in} is independently and identically distributed Extreme Value Type I, the probability that decision maker n chooses alternative i takes the well-known logit formula given by Eq. 2.

$$U_{in} = V_{in} + \varepsilon_{in} \text{ where } V_{in} = \sum_{m} \beta_m x_{imn} \tag{1}$$

$$P_{in} = \frac{e^{Vin}}{\sum_{j \in C} e^{V_{jn}}} \tag{2}$$

Since our focus is on predicting utility levels, rather than hypothesis testing, we adopt an iterative model specification approach. That is, we started with the complete set of street-level components of Table 1 (i.e. the attributes) and then iteratively removed those that were statistically insignificant, where in each step, we removed the most insignificant one and continued this stepwise approach until all parameters were below the threshold level of significance of $\alpha = 0.10$.[2] In addition, to improve model parsimony, we merged the classes 'Building' and 'House' into a single class (which we still named 'Building') and 'Plants' and 'Shrubs' into a single class (still named 'Plants') as they were frequently confused by the segmentation models and exhibited similar effects on utility.

[2] At the start of this process, we fixed the unsegmented pixels to zero for normalisation.

2.2. Application of the residential location choice model

After the development of the residential location choice model, the next steps involve applying the estimated residential location choice model to obtain the spatial distribution of utility derived from street-level conditions for the city of Rotterdam. Importantly, unlike e.g. agent-based simulations or residential sorting studies, our aim is not to predict individual household residential location decisions, but rather to map the distribution of the utility derived from the street-level conditions from a residential location choice perspective, at a city-wide scale.

2.2.1. Step 2A: image data collection for Rotterdam

This step of the method involves the collection of URLs to street-level images of residential areas in the city of Rotterdam. To compile this dataset, we took the following steps:

1. We create a grid of points with 100-metre spacing within areas designated as residential areas within Rotterdam.
2. We identify the IDs of 360-degree images from Google's street view database, taken in the last decade (i.e. year $\geq$2014), following the process described in (Garrido-Valenzuela et al., 2023).
3. From each image ID, we generated two image URLs, with 90-degree angles to the direction of the street (to both directions). This latter ensures that the images are 'window views', in line with the orientation of the images used by Van Cranenburgh and Garrido-Valenzuela (2025) in their stated choice experiment.
4. The municipality of Rotterdam has various suburbs and a large stretch of land area primarily designated for port activities. We excluded images from suburbs outside the city's core and neighbourhoods with primarily port activities. Additionally, we dropped images taken on highways and primary roads.
5. The final database contains URLs from over 200 thousand street-level images of residential streets in Rotterdam.

2.2.2. Step 2B: image segmentation

We segment the 200k geotagged images of Rotterdam from step 1A. Then, we use the foundation segmentation model with the same input prompts as used for developing the residential location choice model (see Table 2). Compared to conventional task-specific segmentation models, a drawback of the foundation segmentation model is that it is considerably more computationally demanding – approximately twenty times more so in our setup. Consequently, the choice of segmentation model should be guided by the objective and scale of the application: task-specific models are preferable when computational efficiency is paramount (e.g., when mapping large cities on a regular basis), whereas foundation models are advantageous when the detection of novel or fine-grained classes is essential.

2.2.3. Step 2C: model application

Next, we utilise the estimated residential location discrete model to compute utilities derived from the street-level conditions based on the segmented images of Rotterdam. Notably, in the application of the model, we exclude utilities associated with the numeric attributes 'Commute travel time' and 'Housing cost', which were part of the estimation data but are not relevant to our application. Excluding the numeric attributes thus allows us to isolate the utility levels that reflect only the street-level conditions.

2.2.4. Step 2D: aggregation

The final step of the method involves the spatial aggregation of the results. We spatially average the utilities by postcode level 5, henceforth called PC5. In the Netherlands, the postcodes consist of four digits followed by two letters (e.g., 1234 AB). The PC5 level (e.g. 1234A) typically corresponds to a small neighbourhood, with an average area equivalent to approximately 10.7 hectare. This level of aggregation provides a good balance between having a sufficient number of images

Fig. 3. Comparison between the street-level components extracted using a conventional task-specific segmentation model (Mask2Former) and a foundation segmentation model (GroundingDINO + SAM) (image source: Google).

per spatial unit and maintaining a high resolution in the spatial map. The area of study in Rotterdam comprises 863 PC5 areas, with a median of 144 images per area. As a consequence, the visualised results are presented at a slightly higher spatial aggregation level than the individual street segments from which the street-level conditions were derived.

3. Results

This section presents the results and comprises two parts. Section 3.1 presents the result of the residential location choice model development; Section 3.2 presents the substantive results of the application of the model to the city of Rotterdam.

3.1. Residential location choice model results

Table 2 presents the estimation results of four residential location discrete choice models. The first two columns report models previously introduced in the paper where the dataset was originally presented. Specifically, the first column shows the results of the Computer Vision-enriched DCM, an end-to-end model that directly learns from image data without relying on separate pre-trained segmentation models (cf. Nathvani et al., 2023; Weichenthal, Hatzopoulou, & Brauer, 2019 for a

Table 2
Estimation results.

		Benchmark models				*Segmentation-based models*			
Data set	**Model**	**CV-DCM**		**Basic RUM-MNL**		**Mask2former Cityscapes**		**GroundingDINO + Segment Anything**	
	No. parameters	86 m		2		11		16	
Estimation dataset	Log-likelihood	−5724.0		−5953.6		−5680.5		−5664.8	
$N = 9783$	ρ^2	0.156		0.122		0.162		0.165	
	Cross entropy	0.585		0.609		0.581		0.579	
	Log-likelihood	−1137.6		−1193.7		−1166.7		−1154.8	
Hold-out dataset	ρ^2	0.158		0.116		0.136		0.145	
$N = 1948$	Cross entropy	0.585		0.613		0.599		0.593	
		Est	**tval**	**Est**	**tval**	**Est**	**tval**	**Est**	**tval**
Numeric attributes									
	β_{hhcost}	−0.96	−38.40	−0.863	−35.56	−0.918	−35.96	−0.921	−36.03
	β_{tt}	−0.24	−9.23	−0.208	−8.30	−0.231	−8.93	−0.229	−8.88
Visual components									
	Proportions								
	$\beta_{unsegmented}$					0.000	–fixed	0.000	–fixed
	β_{road}					0.000	–fixed	0.000	–fixed
	β_{fence}					0.000	–fixed	0.000	–fixed
	β_{car}					−0.834	−3.30	−0.715	−2.48
	β_{truck}					−2.909	−2.39	−1.329	−2.25
	$\beta_{building}$							0.996	4.63
	$\beta_{sidewalk}$							0.610	2.56
	β_{sky}					1.306	8.99	2.435	8.01
	$\beta_{terrain}$					1.091	7.66		
	$\beta_{vegetation}$					1.205	14.54		
	β_{trees}							2.355	8.84
	β_{plants}							1.767	7.82
	β_{grass}							1.490	7.85
	β_{water}							1.855	3.78
	β_{pole}							−8.021	−2.18
	$\beta_{transformer/bin}$							−5.894	−3.85
	Counts								
	$\beta_{motorcycle}$					−0.181	−2.31		
	$\beta_{traffic\ light}$					−0.339	−3.05	−0.107	−1.88
	$\beta_{traffic\ sign}$					−0.175	−4.55	−0.138	−3.12
	$\beta_{bicycle}$							−0.031	−1.94
	β_{truck}					−0.173	−1.98		

discussion on end-to-end vs feature-based approaches). The second column contains the results of a basic MNL model that omits street-level conditions entirely, serving as a lower bound. Columns 3 and 4 report the estimation results of the two residential location discrete choice models developed in this study using image segmentation; the former using the Mask2Former segmentations, the latter using the foundation model segmentations.

We begin by examining the model fits, from which several observations can be drawn. First, focusing on performance on the hold-out set, the two segmentation-based models perform between the CV-DCM and the basic RUM-MNL benchmark. This is consistent with expectations: we expect the segmentation-based models to outperform the basic MNL, which omits street-level conditions entirely, but to fall short of the CV-DCM, which can capture complex interactions between street-level components and account for nonlinearities –capabilities the segmentation-based linear-additive models lack. The segmentation models attain ρ^2 values of about 0.14, which lies within the expected range for stated choice experiments (typically 0.1–0.4, see Louviere, Hensher, & Swait, 2000). It is therefore important to note that these modest ρ^2 values should not be interpreted as indicating a poor model. In fact, well-designed choice experiments deliberately feature nontrivial trade-offs that make the data maximally informative for parameter estimation (Kanninen, 2002; Rose & Bliemer, 2009). Furthermore, for the segmentation-based models, a small gap is noticeable between the ρ^2 values on the estimation and hold-out sets, likely due to the iterative estimation procedure. This minor discrepancy remains within acceptable bounds and demonstrates that the developed models generalise well to unseen data.

When comparing the two segmentation-based models, we find that the model employing the foundation segmentation approach achieves a better fit on the hold-out set, with a log-likelihood of −1154.8 compared to −1166.7 for the conventional segmentation model. Although this improvement, e.g. in terms of in ρ^2, appears modest, the difference is statistically significant and meaningful within the discrete choice modelling context, as even small increases in model fit can reflect nontrivial improvements in explanatory performance (Ben-Akiva & Lerman, 1985). Moreover, the foundation-based model yields five additional statistically significant parameters. This suggests that the more refined segmentations produced by the foundation model allow for the inclusion of additional street-level components that influence residential location choices. To rule out the possibility that the improvement in model fit is solely attributable to more accurate segmentation –rather than to the refinement and extension of class labels– we estimated a model using the foundation-based segmentation constrained to the original Cityscapes classes. The resulting model shows an improvement in fit (−1158.8 vs. –1166.7, see Appendix A), but does not exceed the model fit of the model with the refined classes. This result thus lends support to the notion that both increased segmentation accuracy and the use of more granular classes contribute to the improved explanatory power of the foundation-based model. Nonetheless, further research is needed to explore both explanations in greater depth.

Next, we look at the parameters of the segmentation-based models. Firstly, in line with expectations, we see that the estimates associated with the numeric attributes, i.e. for housing cost and commute travel time, have the expected negative signs. Moreover, the ratios of $\beta_{tt}/\beta_{hhcost}$ are virtually the same as those of the benchmark models. This was to be expected as the experiment is designed in such a way that no correlations exist between the numeric attributes and the images that were presented, and the feature extraction and model estimation are fully decoupled. Therefore, we can be confident that preferences for housing

cost and travel time are accurately identified and are not inadvertently influenced by visual cues related to, e.g. the socioeconomic conditions, such as income levels, property prices, or urban density (Law et al., 2019; Suel, Bhatt, Brauer, Flaxman, & Ezzati, 2021), that might correlate with housing cost.

Secondly, the parameters associated with the street-level components also all exhibit the intuitively expected signs. When interpreting the parameters associated with proportional components, it is crucial to consider the reference point, which is defined by the fixed (i.e. statistically insignificant) components. Although the specific set of insignificant components varies between the two models, both share two relatively abundant reference ones, namely 'Road' and 'Fence'. As a result, the effective reference point is largely consistent across the two models, allowing for a meaningful comparison of the estimated parameters. Considering the reference, positive coefficients, such as those associated with 'Vegetation' and 'Sky', imply that increasing the presence of these features, at the expense of road or fence coverage, increases the utility derived from street-level conditions significantly. Conversely, negative coefficients –such as those for 'Cars' and 'Trucks'– indicate that a higher proportion of these elements, relative to the reference classes, is associated with significantly reduced utility. Furthermore, insignificant parameters should not be interpreted as residents not caring about the component; instead, it means that they are indifferent between the component and the reference components. Thirdly, the analysis reveals nuanced differences among vegetation types. In particular, the parameter estimates associated with 'Trees' and 'Plants' are considerably higher than those for 'Grass', indicating that individuals place greater value on the presence of trees and plants compared to grass on a per-pixel basis. This finding highlights the importance of distinguishing between different vegetation types when studying people's perceptions and preferences of the urban space.

Finally, in Fig. 4, we explore the utility predictions derived from the street-level components for the three images presented in Fig. 3 using our best model. This examination aims to provide deeper insights into the model's predictions. Specifically, the part-worth utilities for each street-level component are shown in the images. To aid interpretation, we use a colour scale, where greener implies greater positive utility and deeper red implies greater disutility. In general, the part-worth utilities look plausible. A particularly notable pattern is the relatively large negative impact of transformer houses. In conclusion, the combination of intuitively reasonable parameter signs and the interpretability of the visual outputs in Fig. 4 reinforces the view that the developed residential location choice model provides a credible representation of residential preferences. In light of these results, we use the foundation-based segmentation model for our application.

3.2. Residential location choice model application

Before showing the main result, i.e. the spatial distribution of the utility derived from residential street-level conditions in Rotterdam, this subsection starts by examining the joint distributions of the street-level components across space (Section 3.2.1). Understanding their multivariate nature helps to interpret the spatial patterns of street-level conditions better. Then, Section 3.2.2 presents the main result. Section 3.2.3 delves deeper into the results by investigating specific areas in more detail that are characterised by the greatest amount of a component. Finally, Section 3.3 compares our findings with those obtained using traditional approaches, highlighting how this study provides complementary insights.

3.2.1. Joint distributions of semantic components

Fig. 5 shows the bivariate relationships between the nine most important[3] street-level components aggregated at the PC5 level. We refrained from depicting the remaining five significant components as this would render the figure too small. We add 'Road' and the (total) utility derived from the street-level conditions to complete the picture. Each scatter plot shows the pairwise relationships between these components, with histograms on the diagonal representing their individual distributions.

Fig. 5 provides the following insights. Firstly, the observed correlations between the street-level components are intuitive. For instance, grass and trees are positively correlated, while buildings and grass are negatively correlated. Secondly, the correlations between the components are not excessively high ($\rho \leq 0.90$). The highest correlation is between 'Building' and 'Trees' ($\rho = -0.83$). It is important that correlations are not too high to ensure that the individual effects of these components can be disentangled in the model. Thirdly, histograms on the diagonal reveal that the street-level components are predominantly normally distributed. Also, the distribution of the utility derived from street-level conditions (shown in the bottom-right corner) follows the same pattern. Right-skewed distributions are seen for 'Grass', 'Plants', 'Transformer/waste bin' and 'Water'. Utility values range from 0.49 to 1.81, with a mean of 0.99. To interpret these numbers, it is important to note that utility does not have an absolute scale. In RUM-based discrete choice models, only differences in utility are meaningful (Train, 2003). In other words, we cannot interpret the sign of the utility level itself, only its relative position within the distribution. For example, a utility level above 0.99 indicates that the street-level conditions in that PC5 area are more attractive than average.

Furthermore, the bivariate relationship involving the component 'Road' warrants additional attention. Fig. 5 shows that 'Road' hardly correlates with utility ($\rho = 0.01$). This low correlation may initially appear intuitive, given that β_{road} is fixed to zero in the estimation of the discrete choice model. However, the relationship is more complex due to the compositional nature of street-level components: an increase in the proportion of road necessarily reduces the proportion of other elements. Fig. 5 indicates negative correlations of road proportion with 'Building' and 'Plants', while road proportion correlates positively with 'Trees' and 'Sky'. Thus, the low overall correlation of road with utility arises because road proportion simultaneously associates with elements having both positive and negative marginal utilities, effectively neutralising the net impact on overall utility.

3.2.2. Spatial distribution of residential street-level condition

The middle plot of Fig. 6 shows the main result of this study: the spatial distribution of the utility derived from street-level conditions to residential location choices in Rotterdam. The colour scale is such that the colour red indicates a comparatively low utility derived from the residential street-level conditions, and the colour green indicates a comparatively high utility derived from the residential street-level conditions. Fig. 6 reveals several key insights into the distribution of the street-level conditions in Rotterdam.

Firstly, the city centre (encircled in cyan), which comprises the neighbourhoods 'CS-Kwartier', 'Stadsdriehoek', 'Cool', 'Oude Westen', 'Dijkzigt', 'Nieuwe Werk', exhibits comparatively poor street-level conditions. Except for 'Dijkzigt' and 'Nieuwe Werk', they have below-average street-level conditions. This suggests that the high real-estate prices typically associated with central locations cannot be attributed to the attractiveness of street-level conditions. It indicates that factors other than the immediate street-level conditions, such as proximity to amenities, employment opportunities, or connectivity, play a more significant role in driving up property values in the city centre. The comparative attractiveness of the residential street-level conditions in 'Dijkzigt' and 'Nieuwe Werk' is explained by the presence of the museum

[3] The importance was determined by computing the part-worth utility for each component and sorting them based on their absolute impact on utility.

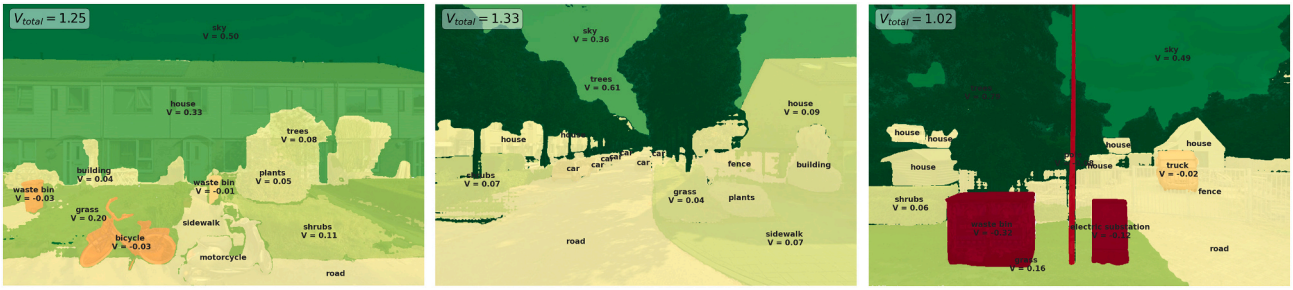


Fig. 4. Visualisation of part-worth utility derived from street-level components using GroundingDINO + SAM.

district and a medium-sized city park.

Secondly, the best residential street-level conditions are found on the city's edges, located particularly near parks and green areas. This spatial pattern reflects the city's spatial hierarchy, where the older, high-density areas dominate the core, which is surrounded by more suburban or commercial/industrial zones. Notable examples include areas near 'Kralingse Bos' in the North and 'Charlois Zuidrand' in the South, as well as residential neighbourhoods with abundant greenery, including trees and plants, such as 's-Gravenland' in the East. These findings underline the importance of greenery in contributing to the attractiveness and utility derived from street-level conditions for the residential location choice.

Thirdly, street-level conditions vary considerably, even within small areas. For example, in the neighbourhoods 'Hillegersberg Noord', 'Schiebroek' and 'Molenlaankwartier' in the Northern part of the city, (dark) red areas stand out in an otherwise attractive (green) area, illustrating how conditions can rapidly improve or deteriorate even within close proximity. So, even though the Northern part of the city is generally considered an upscale and attractive neighbourhood, this area contains pockets where the street-level conditions are not appealing..

Fourthly, and perhaps most surprisingly, the southern neighbourhoods of Rotterdam perform moderately well in terms of residential street-level conditions. In common parlance, the southern neighbourhoods, such as 'Bloemhof', 'Tarwewijk', and 'Pendrecht', are perceived as 'probleemwijken' (problem areas) due to higher poverty and crime rates. However, these results show that despite the challenges in these neighbourhoods, the street-level conditions there are not as poor as one might expect. In other words, there are positive aspects to the street-level conditions in these neighbourhoods that may not be immediately apparent in conventional socioeconomic analyses.

3.2.3. *Areas of interest*

Lastly, we investigate the utility contribution of the street-level components for specific areas. We focus on the six blue-encircled PC5 areas shown in Fig. 6. These areas are selected because they feature the highest proportion of six relevant street-level components and, therefore, make interesting cases. To provide the reader with a better grasp of these areas, collages with random samples of street-level images taken from these areas can be found in Appendix B. Additionally, Appendix C shows the locations of the street-level components for these areas within the distribution of the city as a whole (at the PC5 level).

The bar plots on both sides of Fig. 6 show how these areas deviate from the mean utility stemming from that component for the city as a whole. These deviations are computed by comparing the mean utility derived from each semantic component for the city as a whole with that of the area under investigation. The bar outlined in orange highlights the component for which this PC5 area shows the greatest proportion. For instance, PC5 area '3039E', in neighbourhood 'Blijdorp', has the highest proportion of cars, while PC5 area '3011V' in 'Stadsdriehoek' features the highest proportion of water. One exception to the highlighting, is PC5 area '3026A' in the neighbourhood 'Tussendijken' (located in the Western part). This area features the highest proportion of 'Road'. However, since the proportion of road was fixed to zero in the estimation (and thus is the reference), this component is not displayed in the bars. The effect of more roads is thus indirectly captured via the decrease in other components (see the last paragraph of Section 3.1 for a discussion on this matter). Finally, in the plot titles, also the mean utility derived from the street-level conditions (V_{img}^{mean}) and the deviation of the street-level utility from the mean (ΔV) are reported.

The bar plots provide several insights into the variation in street-level conditions in Rotterdam. Firstly, we see that two areas under investigation: '3039E' ('Blijdorp') and '3011P' ('Stadsdriehoek'), achieve a by and large similar street-level utility of $V_{img}^{mean} \approx -0.18$. Interestingly, the bar plots show that different street-level components explain this. Area '3039E' features an above-average proportion of trees, which contribute positively to the street-level conditions. However, the area modestly underperforms in terms of all other considered components, reducing its attractiveness. In contrast, area '3011P' especially lacks trees, making the area comparatively less attractive. But it has above-average amounts of buildings and sidewalks, and a below-average presence of cars, which makes the area comparatively more attractive. Looking at images from these areas (in Appendix B) further attests that these areas have distinct visual characters.

Secondly, the area '3065A' in 's-Gravenland' features a highly attractive street-level condition. A glance at the images of this area (in Appendix B) confirms the high street-level conditions. As the bar plot shows, its attractiveness is explained by the above-average presence of trees, sky, grass, plants, and water. The two red bars for 'Building' and 'Sidewalk' suggest that this area features below-average shares of these components, and that increasing them would make the area even more attractive. However, the bars show the utility impact of the components, relative to the city as a whole. Therefore, it does not necessarily reveal how to further enhance the attractiveness of an area. Indeed, if buildings were to replace trees, it would cause a net decrease in street-level attractiveness since $\beta_{trees} > \beta_{building}$. Vice versa, if buildings were to replace cars, it would cause a net increase in street-level attractiveness since $\beta_{car} < \beta_{building}$.

Thirdly, the areas '3026A' and '3012X' are the two least attractive areas (out of the six selected for further investigation). Area '3026A' has the highest share of road; area '3012X' has the highest share of buildings. As can be seen in the bar plots, both areas have a similar utility profile. Both perform comparatively badly in terms of the presence of trees, sky, grass and plants. Only the share of building results in an above-average impact on the street-level conditions. Enhancing the street-level conditions in these areas requires increasing the presence of trees, sky, grass and plants. However, to achieve this requires different actions because of their different visual characters (see the images in Appendix B and the histograms in Appendix C). Area '3026A' features a lot of roads and an average proportion of buildings and sky; area '3012X' features mostly buildings and hardly any roads. To improve the attractiveness of '3026A', the natural course of action would be to replace roads with trees and plants. In contrast, improving street-level conditions in area '3021X' would require reducing the presence of buildings, e.g. by tearing down some buildings to make the sky visible and free up

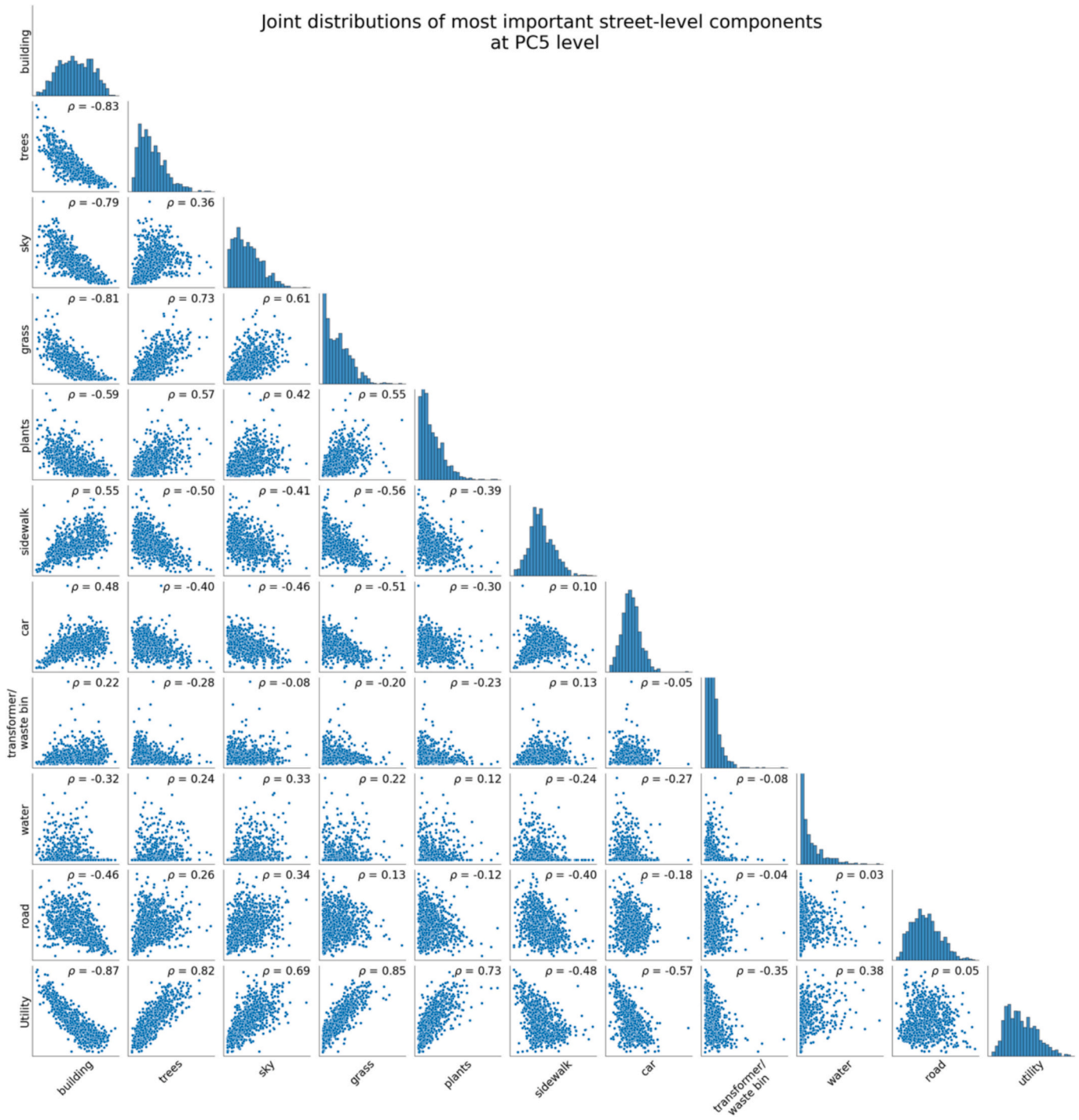


Fig. 5. Joint distribution of the most important street-level components at PC5 level.

space for trees. These two cases illustrate how detailed insights into the local street-level conditions can be used to inform targeted policies to improve the street-level conditions.

3.3. Comparison with related indicators and methods

To thoroughly understand how the utility derived from street-level conditions relates to existing indicators of residential quality, we compare our results with three complementary sources: (i) real-estate prices (CBS, 2024), (ii) Leefbaarometer scores (Mandemakers et al., 2021), and (iii) an image-based perception model trained on PlacePulse 2.0 data to predict perceived beauty (Dubey, Naik, Parikh, Raskar, & Hidalgo, 2016; Zhou, Lapedriza, Khosla, Oliva, & Torralba, 2017). These indicators differ in their underlying concepts and data sources, allowing us to situate our findings within a broader empirical and perceptual context. The result of these comparisons is shown in Fig. 7.

3.3.1. Real-estate prices

The real-estate price data (Waardering Onroerende Zaken, abbreviated as WOZ) from CBS represent official property valuations based on recent transactions. These values capture aggregated market assessments that reflect not only residential quality but also a wide range of additional factors, including location, dwelling characteristics, accessibility, and neighbourhood amenities. At the PC5 level, the correlation

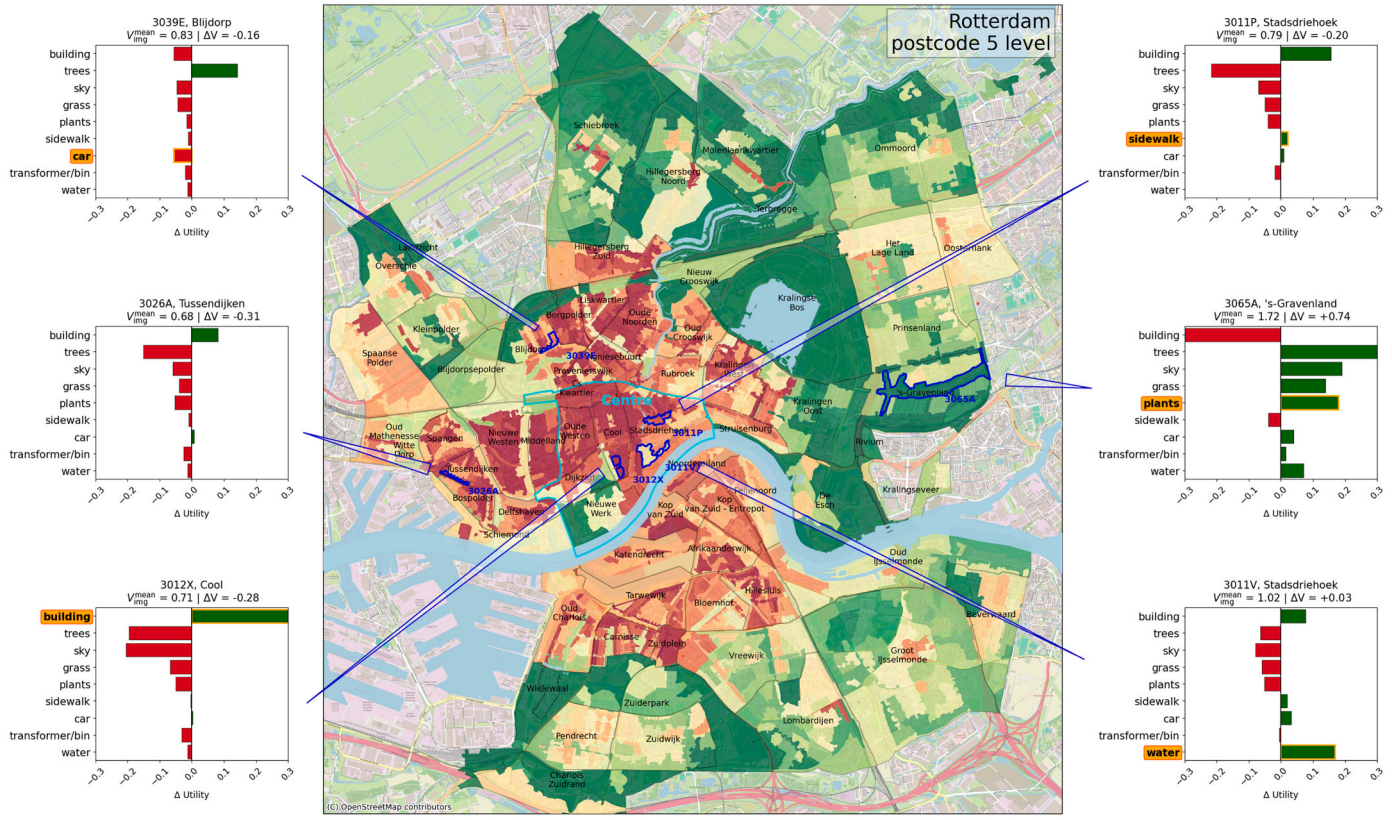


Fig. 6. Spatial distribution of residential street-level conditions utility (from images) in Rotterdam (middle). Locations of areas with the highest proportion of street-level components are encircled in blue. Impact on street-level utility compared to the mean is shown on both sides for selected areas. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

between WOZ and the utility derived from street-level conditions is modest ($\rho = 0.24$). This modest association is largely explained by the city centre, where high real-estate valuations are primarily driven by accessibility and market demand rather than by the immediate quality of the street environment.

3.3.2. *Leefbarometer scores*

The Leefbarometer is a nationwide monitoring instrument developed by the Dutch Ministry of the Interior (BZK) and Atlas Research. It combines a stated preference model, based on the WoonOnderzoek Nederland (WoON 2018) survey (~67,000 respondents), with a hedonic price model estimated from over 700,000 housing transactions (2017–2019). Together, these models capture how physical, social, and environmental characteristics of neighbourhoods jointly shape residents' residential perceptions and valuations. The "physical environment" dimension, used in this comparison, integrates indicators such as green space, air quality, building height, noise, and proximity to infrastructure. At the PC5 level, the correlation between this dimension and our street-level utility is moderate ($\rho = 0.47$). While both indicators capture aspects of environmental quality, the Leefbarometer embeds broader infrastructural and socio-environmental factors, whereas our utility measure isolates the directly observable visual qualities of the residential environment.

3.3.3. *Image-based perception model (PlacePulse 2.0)*

To assess how utility derived from street-level conditions relates to perception-based assessments of the built environment, we apply an image-based model trained on the PlacePulse 2.0 dataset to predict perceived "beauty" from street-level imagery (Dubey et al., 2016; Zhou et al., 2017). We apply this model to the same set of Rotterdam street-level images used in our utility analysis, following the same application pipeline but replacing the residential location choice model with the PlacePulse-based perception model using the ZenSVI implementation (Ito et al., 2025). In line with expectations, we observe that the beauty perception scores correlate strongest with the utility derived from street-level conditions ($\rho = 0.66$). However, the correlation remains moderate, indicating that the two measures capture related but distinct aspects of the urban environment.

Altogether, the comparisons reveal that the utility derived from street-level conditions captures a related yet distinct dimension of the urban environment. To examine this further, we estimated a linear model using the three indicators: real-estate prices, Leefbarometer scores, and PlacePulse beauty perceptions, to predict the street-level utility. The predicted scores correlated only moderately with the observed utility ($\rho = 0.72$), providing further evidence that street-level utility captures additional variation not explained by a simple combination of these established indicators. We therefore conclude that the street-level utility derived from our model offers complementary insights to existing methods and indicators of residential liveability.

4. Conclusion and discussion

This study provides new insights into the spatial distribution of utility derived from street-level conditions from a residential location choice perspective and examines the factors shaping this distribution. Thereby, it advances the residential location choice literature by offering insights into the role street-level conditions play in these decisions. In doing so, it complements previous research on urban environment *perceptions-based* on street-level imagery (e.g. Zhang et al., 2018) by introducing a *preference-based* counterpart. As such, the study offers urban planners fine-grained, spatially detailed insights into the street-level conditions, thereby providing actionable input for localised

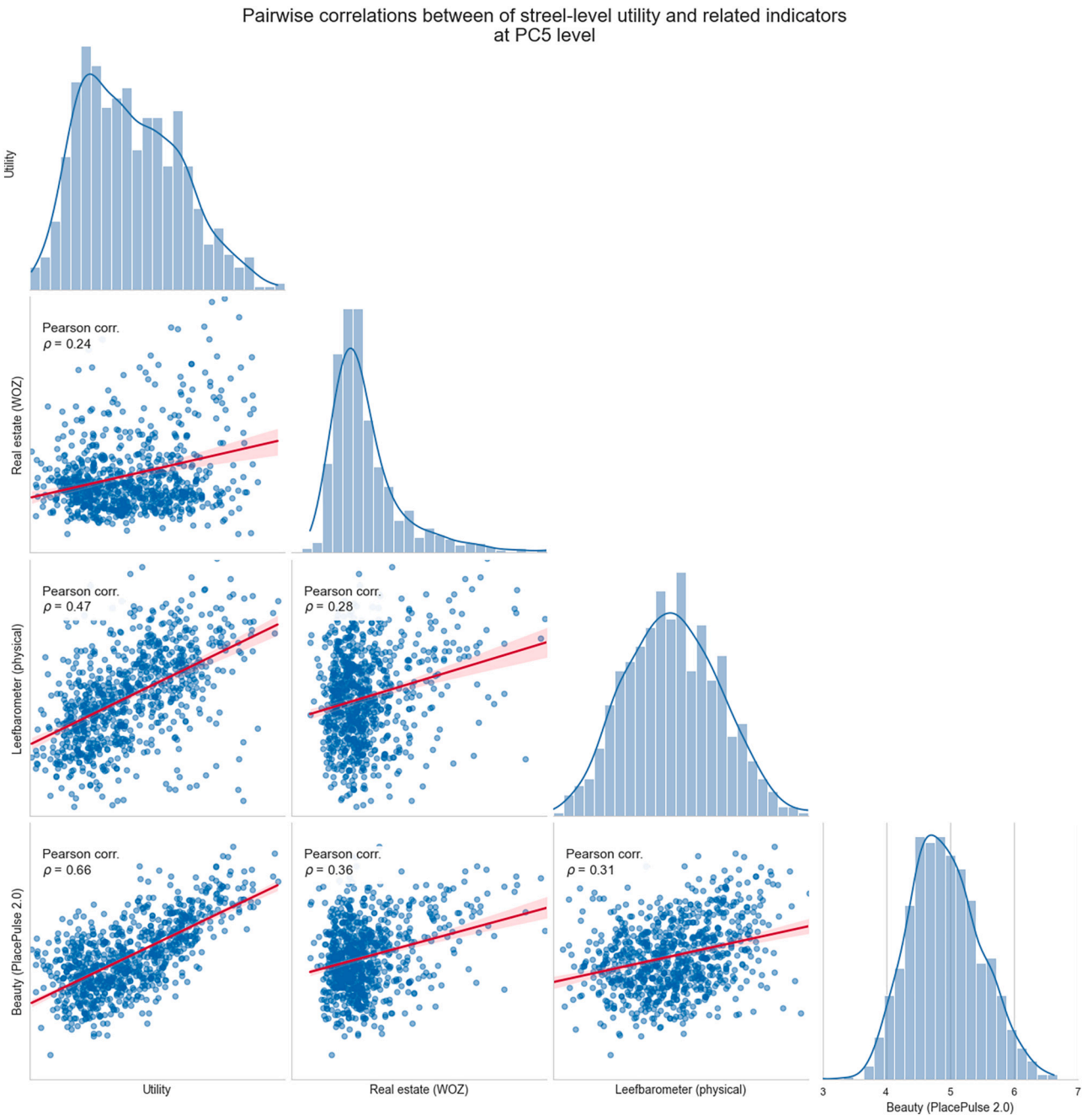


Fig. 7. Comparison of street-level utility with related methods and indicators at PC5 level.

improvements in the built environment. Beyond its substantive insights into residential preferences, this study also makes a methodological contribution. We demonstrate how foundation segmentation models can be integrated into discrete choice modelling to capture fine-grained street-level conditions. In contrast to traditional census-based approaches or conventional segmentation methods with fixed object classes, our pipeline employs prompt-based detection to both refine existing categories (e.g., distinguishing trees, shrubs, and grass) and identify previously unlabelled but policy-relevant objects such as transformer houses. This integration enables the first utility-based, city-wide mapping of street-level conditions, providing a new avenue for urban analysis that combines the solid behavioural foundation and interpretability of DCM with the flexibility of foundation vision models.

Several limitations of this study should be acknowledged, providing avenues for future research. Firstly, the utility levels predicted from our residential location choice model are based on mean population preferences. However, residential location preferences are inherently heterogeneous (Smith & Olaru, 2013). Compounding this complexity, individuals may self-select residential locations that align with their specific preferences, such as street-level conditions (Cao, 2014; Van Wee, 2009). Future studies could further investigate the impact of heterogeneous preferences and self-selection effects, potentially through the use of microsimulation (López-Ospina, Martínez, & Cortés, 2016; Waddell, 2002). This would enable the creation of spatial maps that better reflect how individuals derive utility from the street-level conditions in their living environments.

Secondly, the images used in our analyses all have a perpendicular angle to the street. This orientation may miss certain street-level components that become apparent from a wider or more comprehensive view. For instance, a line of trees might make a perpendicular view feel

'closed', while a parallel or side-angle perspective could convey a sense of openness. Additionally, narrow perpendicular views may obscure Gestalt features, such as the sense of common fate created by a row of trees. Future research could consider conducting stated choice experiments using wide-angle street-level images or full panoramic views to provide a more comprehensive view of the street-level conditions to the respondents, and enabling further refinement of the residential location choice models.

Thirdly, and related to the previous point, while this study focuses on the utility derived from street-level conditions in residential location choice analyses, it does not account for the potential utility derived from the visual appearance of the broader neighbourhood, such as e.g. the walkability of the neighbourhood (Miranda, Fan, Duarte, & Ratti, 2021). The utility people derive from street-level conditions may interact with their perceptions of the wider area. For example, in the case of 'Hillegersberg Noord', a neighbourhood featuring a small area with unattractive street-level conditions, residents might still derive higher-than-expected utility from these conditions due to the overall attractiveness of the larger area, as people can be biased to see what they want to see (Balcetis & Dunning, 2006).

Fourthly, the images used in this study are taken at ground level, meaning that our model predicts utility derived from street-level conditions from this specific perspective. However, in areas with high-rise buildings, residents may also derive utility from views of the skyline, or the openness provided by higher-altitude perspectives. The current study does not consider these factors, but future research could explore the utility residents derive from such views.

Fifthly, related to our methodological contribution, while the extended capabilities of foundation segmentation models allow researchers to detect new object classes, more research is needed to better understand their potential and limitations in urban visual environments. For instance, in our application, transformer houses and waste bins were frequently confused, highlighting the importance of further testing and validation of these models. Rigorous comparative studies are needed to benchmark their performance against conventional approaches. Additionally, fine-tuning such models on domain-specific datasets could improve their ability to distinguish between visually similar objects that are important for understanding urban environments (Kerssies, De Geus, & Dubbelman, 2024).

Finally, a limitation of our study lies in the use of proprietary Google Street View imagery. While this source offers high-quality and geographically consistent visual data, it also introduces potential biases related to spatial coverage, temporal updates, and viewing perspective. For instance, Google's vehicles tend to collect imagery more frequently in densely populated and economically active areas, leading to an underrepresentation of peripheral or less economically developed neighbourhoods, cities, and countries. In addition, Google's terms of service place strict restrictions on the reuse and redistribution, even for academic research. This limits the extent to which datasets can be shared and models can be fully replicated, thereby posing a challenge to the principles of open science and reproducibility. This concern echoes recent calls in the literature (e.g. Danish et al., 2025) to adopt open and FAIR-compliant alternatives such as Mapillary. Future research should consider using open-source imagery to improve reproducibility.

CRediT authorship contribution statement

Sander van Cranenburgh: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Francisco Garrido-Valenzuela:** Writing – review & editing, Methodology, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT-4o in order to improve the readability and language of the manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the TU Delft AI Labs programme and the Urban Intelligence and Liveability project, funded by Convergence-programme AI, Data & Digitalisation and the Municipality of Rotterdam. Furthermore, the authors express gratitude to Lanlan Yan for her valuable contributions to the early development of the paper. The author also would like to thank civil servants at the municipality of Rotterdam, particularly Tommie Perenboom, for their assistance in interpreting the substantive results of the case study.

Appendix A. Estimation results of residential location choice model that uses foundation-based segmentation with the original Cityscape classes

			GroundingDINO + Segment AnythingCityescapes classes only	
Data set	**Model**			
	No. parameters		14	
Estimation dataset $N = 9783$	Log-likelihood		−5700.1	
	ρ^2		0.159	
	Cross entropy		0.583	
	Log-likelihood		−1158.8	
Hold-out dataset $N = 1948$	ρ^2		0.142	
	Cross entropy		0.595	
			Est	**t-val**
Numeric attributes				
		β_{hhcost}	−0.919	−36.04
		β_{tt}	−0.231	−9.01
Visual components				
	Proportions			
		$\beta_{unsegmented}$	0.000	–fixed
		β_{road}	0.000	–fixed

(*continued on next page*)

(*continued*)

Data set	Model		GroundingDINO + Segment AnythingCityscapes classes only	
		β_{fence}	0.000	–fixed
		β_{car}	−1.049	−3.59
		β_{truck}	−1.787	−3.80
		$\beta_{sidewalk}$	0.370	1.76
		β_{sky}	1.013	7.75
		$\beta_{terrain}$	1.450	5.54
		$\beta_{vegetation}$	0.866	8.70
		β_{wall}	−0.740	−3.95
	Counts			
		$\beta_{motorcycle}$	−0.112	−2.81
		$\beta_{traffic\ light}$	−0.128	−3.26
		$\beta_{traffic\ sign}$	−0.086	−4.17
		$\beta_{bicycle}$	−0.035	−2.72
		β_{truck}	−0.104	−3.09

Appendix B. Visual snapshots of areas of interest

Each collage displays two sets of nine images. The left-hand side presents the images with the highest proportion or count of the street-level component under consideration within the PC5 zone. The right-hand side shows a random selection of nine images from the same PC5 zone. At the top left, also the total utility derived from the street-level conditions is shown.

PC5 area '3039E' Blijdorp (highest for component 'Car')

(image source: Google)

PC5 area '3026A' Tussendijken (highest for component 'Road')

(image source: Google)

PC5 area '3012X' Cool (highest for component 'Building')

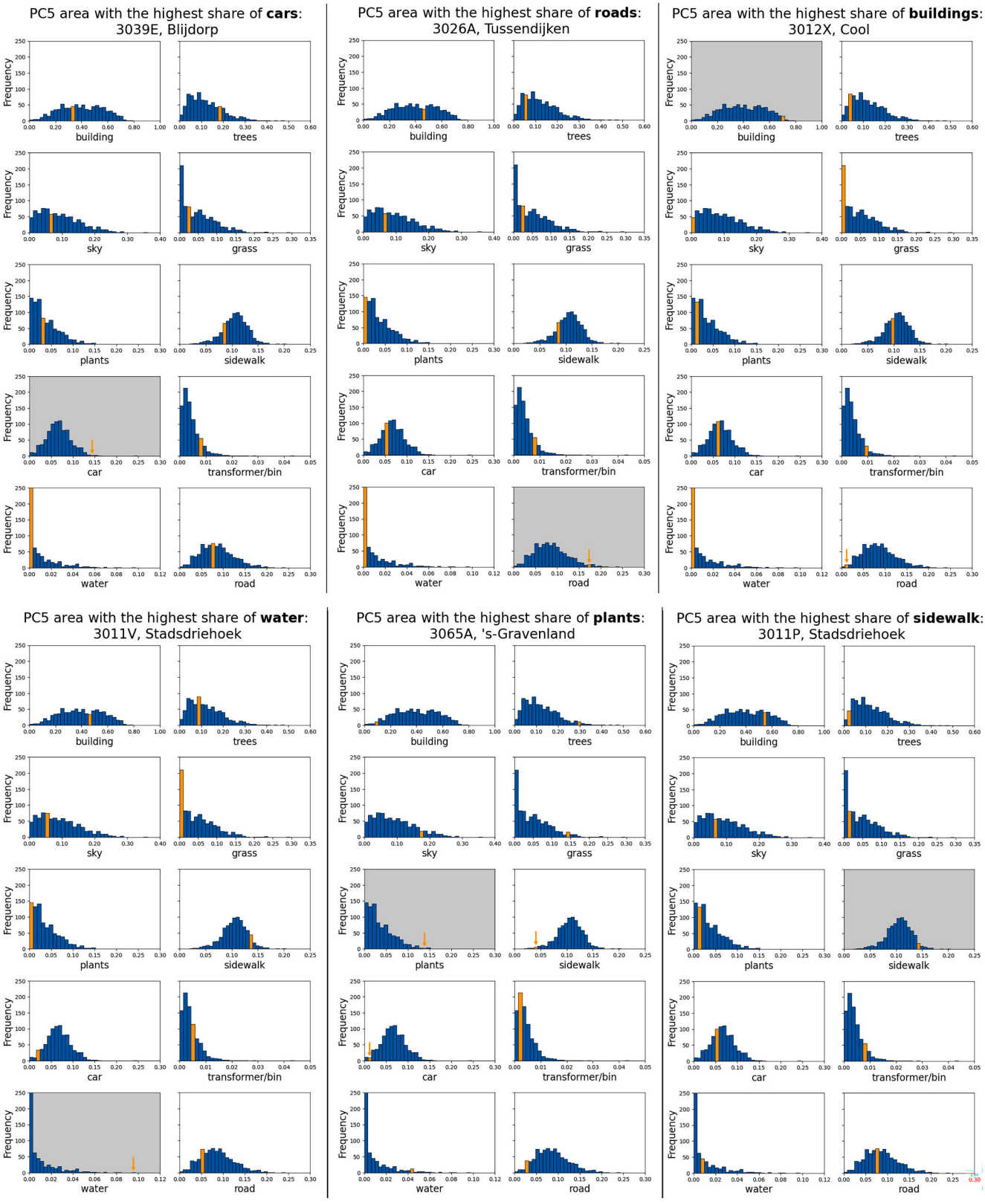
(image source: Google)
PC5 area '3011V' Stadsdriehoek (highest for component 'Water')

(image source: Google)
PC5 area '3065A' 's-Gravenland (highest for component 'Plants')

(image source: Google)
PC5 area '3011P' Stadsdriehoek (highest for component 'Sidewalk')

(image source: Google)

Appendix C. Location of street-level components within the overall distribution


PC5 area with the highest share of **cars**: 3039E, Blijdorp
PC5 area with the highest share of **roads**: 3026A, Tussendijken
PC5 area with the highest share of **buildings**: 3012X, Cool
PC5 area with the highest share of **water**: 3011V, Stadsdriehoek
PC5 area with the highest share of **plants**: 3065A, 's-Gravenland
PC5 area with the highest share of **sidewalk**: 3011P, Stadsdriehoek
Frequency
building
trees
sky
grass
plants
sidewalk
car
transformer/bin
water
road

Data availability

The data and code supporting this study are publicly available at: https://github.com/TUD-CityAI-Lab/Utility-based-street-level-conditions

References

Balcetis, E., & Dunning, D. (2006). See what you want to see: motivational influences on visual perception. *Journal of Personality and Social Psychology, 91*(4), 612.

Ben-Akiva, M., & Lerman, S. R. (1985). *Discrete choice analysis : theory and application to travel demand*. MIT Press.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning, 215*, Article 104217.

Bostanara, M., Siripanich, A., & Rashidi, T. H. (2024). A realistic framework for modelling residential relocation behaviour considering past, present, and future using DDCM: A Sydney case study. *Cities, 153*, Article 105245.

Cao, X. (2014). Examining the impacts of neighborhood design and residential self-selection on active travel: A methodological assessment. *Urban Geography*, 1–20. https://doi.org/10.1080/02723638.2014.956420

CBS. (2024). Waardering onroerende zaken (WOZ), gemiddelde woningwaarde; gemeente, 2024. https://www.cbs.nl/nl-nl/cijfers/detail/85036NED.

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

Clark, W. A., & Ledwith, V. (2007). How much does income matter in neighborhood choice? *Population Research and Policy Review, 26*(2), 145–161.

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., … Schiele, B. (2016). *The cityscapes dataset for semantic urban scene understanding* (Proceedings of the IEEE conference on computer vision and pattern recognition).

Cox, T., & Hurtubia, R. (2021). Latent segmentation of urban space through residential location choice. *Networks and Spatial Economics, 21*, 199–228.

Custers, G., & Willems, J. J. (2024). Rotterdam in the 21st century: From 'sick man'to 'capital of cool'. *Cities, 150*, Article 105009.

Danish, M., Labib, S. M., Ricker, B., & Helbich, M. (2025). A citizen science toolkit to collect human perceptions of urban environments using open street view images. *Computers, Environment and Urban Systems, 116*, Article 102207.

Doucet, B., & Koenders, D. (2018). 'At least it's not a ghetto anymore': Experiencing gentrification and 'false choice urbanism'in Rotterdam's Afrikaanderwijk. *Urban Studies, 55*(16), 3631–3649.

Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). Deep Learning the City: Quantifying Urban Perception at a Global Scale. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision – ECCV 2016 Cham*.

Frenkel, A., Bendit, E., & Kaplan, S. (2013). Residential location choice of knowledge-workers: The role of amenities, workplace and lifestyle. *Cities, 35*, 33–41.

Garrido-Valenzuela, F. (2026). *Pixels, people, places: Computer vision and image embeddings for perception-aware urban analytics* [Doctoral dissertation]. Delft University of Technology https://repository.tudelft.nl/record/uuid:58c5850b-9f9c-4e50-b6d5-2c01c68b9ed2.

Garrido-Valenzuela, F., Cats, O., & van Cranenburgh, S. (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, 102*, Article 101971.

Giles-Corti, B., Bull, F., Knuiman, M., McCormack, G., Van Niel, K., Timperio, A., Christian, H., Foster, S., Divitini, M., & Middleton, N. (2013). The influence of urban design on neighbourhood walking following residential relocation: longitudinal results from the RESIDE study. *Social Science & Medicine, 77*, 20–30.

Gong, X., van den Berg, P., Dane, G. Z., & Arentze, T. (2025). Parental perception of child friendliness and its impact on residential location choice: A stated choice experiment. *Cities, 161*, Article 105914.

Gu, X., Xu, W., Gong, C., & Liu, X. (2025). City centers really lived up to the hype? Evidence from human perceptions of over 4000 communities in China. *Cities, 166*, Article 106278.

Guevara, C., & Ben-Akiva, M. (2006). Endogeneity in residential location choice models. *Transportation Research Record: Journal of the Transportation Research Board, 1977*(1), 60–66. https://doi.org/10.3141/1977-10

Hess, S., & van Cranenburgh, S. (2025). Flexibility without foresight: The predictive limitations of mixture models. *arXiv preprint arXiv:2510.09185*.

Hunt, J. D. (2010). Stated preference examination of factors influencing residential attraction. In *Residential location choice: Models and applications* (pp. 21–59). Springer.

Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., Zhao, T., Ma, R., Fujiwara, K., & Ouyang, J. (2025). ZenSVI: An open-source software for the integrated acquisition, processing and analysis of street view imagery towards scalable urban science. *Computers, Environment and Urban Systems, 119*, Article 102283.

Kanninen, B. J. (2002). Optimal Design for Multinomial Choice Experiments. *Journal of Marketing Research, 39*(2), 214–227. https://doi.org/10.1509/jmkr.39.2.214.19080

Kerssies, T., De Geus, D., & Dubbelman, G. (2024). *How to Benchmark Vision Foundation Models for Semantic Segmentation?* (Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition).

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., … Lo, W.-Y. (2023). Segment anything. *arXiv preprint arXiv:2304.02643*.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy, 74*(2), 132–157. https://doi.org/10.2307/1828835

Law, S., Paige, B., & Russell, C. (2019). Take a look around: using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology (TIST), 10*(5), 1–19.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., … Yang, J. Zhu (2023). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *Europeanconference on computer vision* (pp. 38–55). Springer Nature Switzerland: Cham.

López-Ospina, H. A., Martínez, F. J., & Cortés, C. E. (2016). Microeconomic model of residential location incorporating life cycle and social expectations. *Computers, Environment and Urban Systems, 55*, 33–43. https://doi.org/10.1016/j.compenvurbsys.2015.09.008

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods : analysis and applications*. Cambridge University Press.

Luce, R. D. (1959). *Individual choice behavior; a theoretical analysis*. Wiley.

Mandemakers, J., Leidelmeijer, K., Burema, F., Halbersma, R., Middeldorp, M., & Veldkamp, J. (2021). *Leefbaarometer 3.0*.

McFadden. (1977). *Modelling the choice of residential location*.

Miranda, A. S., Fan, Z., Duarte, F., & Ratti, C. (2021). Desirable streets: Using deviations in pedestrian trajectories to measure the value of the built environment. *Computers, Environment and Urban Systems, 86*, Article 101563.

Nathvani, R., Vishwanath, D., Clark, S. N., Alli, A. S., Muller, E., Coste, H., … Baah, S. (2023). Beyond here and now: Evaluating pollution estimation across space and time from street view images with deep learning. *Science of the Total Environment, 903*, Article 166168.

Pagliara, F., & Wilson, A. (2010). The state-of-the-art in building residential location models. In *Residential location choice: Models and applications* (pp. 1–20). Springer.

Pérez, P. E., Martínez, F. J., & Ortúzar, J.d. D. (2003). Microeconomic formulation and estimation of a residential location choice model: Implications for the value of time. *Journal of Regional Science, 43*(4), 771–789.

Pinjari, A. R., Pendyala, R. M., Bhat, C. R., & Waddell, P. A. (2007). Modeling residential sorting effects to understand the impact of the built environment on commute mode choice. *Transportation, 34*(5), 557–573.

Rose, J. M., & Bliemer, M. C. J. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews: A Transnational Transdisciplinary Journal, 29*(5), 587–617. http://www.informaworld.com/ https://doi.org/10.1080/01441640902827623.

Rui, J. (2023). Measuring streetscape perceptions from driveways and sidewalks to inform pedestrian-oriented street renewal in Düsseldorf. *Cities, 141*, Article 104472.

Samuelson, P. A. (1948). *Foundations of economic analysis*.

Schirmer, P. M., Van Eggermond, M. A., & Axhausen, K. W. (2014). The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use, 7*(2), 3–21.

Seiler, M., Madhavan, P., & Liechty, M. (2012). Toward an understanding of real estate homebuyer internet search behavior: an application of ocular tracking technology. *Journal of Real Estate Research, 34*(2), 211–242.

Smith, B., & Olaru, D. (2013). Lifecycle stages and residential location choice in the presence of latent preference heterogeneity. *Environment and Planning A, 45*(10), 2495–2514.

Suel, E., Bhatt, S., Brauer, M., Flaxman, S., & Ezzati, M. (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment, 257*, Article 112339.

Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

Van Cranenburgh, S., & Garrido-Valenzuela, F. (2025). Computer vision-enriched discrete choice models, with an application to residential location choice. *Transportation Research Part A, Policy and Practice, 192*.

Van Wee, B. (2009). Self-Selection: A Key to a Better Understanding of Location Choices, Travel Behaviour and Transport Externalities? *Transport Reviews, 29*(3), 279–292. https://doi.org/10.1080/01441640902752961

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation, and environmental planning. *Journal of the American Planning Association, 68*(3), 297–314.

Waddell, P. (2006). *Reconciling household residential location choices and neighborhood dynamics. Under revision, sociological methods and research*.

Weichenthal, S., Hatzopoulou, M., & Brauer, M. (2019). A picture tells a thousand… exposures: opportunities and challenges of deep learning image analyses in exposure science and environmental epidemiology. *Environment International, 122*, 3–10.

Xue, F., & Yao, E. (2022). Adopting a random forest approach to model household residential relocation behavior. *Cities, 125*, Article 103625.

Yang, L., Ding, C., Ju, Y., & Yu, B. (2021). Driving as a commuting travel mode choice of car owners in urban China: Roles of the built environment. *Cities, 112*, Article 103114.

Zhang, F., Salazar-Miranda, A., Duarte, F., Vale, L., Hack, G., Chen, M., Liu, Y., Batty, M., & Ratti, C. (2024). Urban Visual Intelligence: Studying Cities with Artificial Intelligence and Street-Level Imagery. *Annals of the American Association of Geographers, 114*(5), 876–897.

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

Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., … Shum, H.-Y. (2022). *Dino: Detr with improved denoising anchor boxes for end-to-end object detection*. arXiv preprint arXiv: 2203.03605.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(6), 1452–1464.

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A., & Torralba, A. (2017). *Scene parsing through ade20k dataset. (Proceedings of the IEEE conference on computer vision and pattern recognition)*.