A utility-based spatial analysis of residential street-level conditions

A case study of Rotterdam

Sander van Cranenburgh[1,*]
Francisco Garrido-Valenzuela[1]

[1]CityAI lab, Transport and Logistics Group,
Delft University of Technology
The Netherlands
*Corresponding author

## 1.	Introduction

Residential location choices shape the infrastructure and functionality of cities. Specifically, individuals and households' decisions about where to live 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 Discrete Choice Models (DCMs) (Guevara & Ben-Akiva, 2006; Hunt, 2010; McFadden, 1977; Pérez et al., 2003). DCMs are grounded in random utility theory, where individuals are assumed to choose the alternative that maximises their utility from a set of discrete alternatives, each conceived as a bundle of attributes (e.g. cost, time, quality) (Ben-Akiva & Lerman, 1985). Residential location choices entail trade-offs along various factors, including (1) *Travel and accessibility-related factors*, such as the commute mode, commuting time, and distances to amenities like schools, stores, hospitals and playgrounds; (2) *Socioeconomic environments*, such as income levels, ethnic distribution, age and education level; and, (3) *Built environments and street-level conditions*, such as built density, housing types and typology, street layout, traffic safeness, greenness, parking conditions, and disorders (e.g. due to litter, graffiti, or weeds) (Giles-Corti et al., 2013).

Despite the acknowledged importance of the built environment and street-level conditions to residential location choices, they are commonly neglected when analysing them (Schirmer et al., 2014). This neglect relates to data practices. Residential location choice models have traditionally been built using census data, which does not contain detailed information about the built environment and street-level conditions. Only a limited number of residential location choice models have incorporated built environment variables, and those have typically relied on (tabular) cadastral data (Pinjari et al., 2007). While useful, these data often fall short of capturing detailed street-level conditions, such as the number of cars driving or parked in the street, the presence of sidewalks, or the amount and type of trees. Additionally, they lack information on more intangible street-level conditions, such as 'openness', 'scale, 'order', and 'continuity', all of which are known to influence how people perceive and evaluate streetscapes (Gjerde, 2011). As a result, comparatively little is known about how local street-level

conditions – from a residential location choice perspective – are spatially distributed within cities.

Unlike tabular 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. Importantly, nowadays, images containing information on street-level conditions are widely available from tech firms like Google, Apple, and Baidu's map services. Since these images became widely accessible, researchers have extensively utilised them to analyse and understand urban environments (Zhang et al., 2024). Notably, the pioneering work of Dubey et al. (2016) and subsequent studies have significantly advanced our understanding of how urban spaces are perceived in terms of e.g. safeness, vibrancy and liveliness using street-level images (Liu et al., 2017; Ma et al., 2021; Wei et al., 2022; Zhang et al., 2022; Zhang et al., 2018).

In addition to these developments, a new type of discrete choice model has recently been proposed, called Computer Vision enriched Discrete Choice Models (henceforth abbreviated as CV-DCMs) (Van Cranenburgh & Garrido-Valenzuela, 2025). CV-DCMs can handle images by integrating traditional random utility theory-based choice models and computer vision models. And, because the CV-DCM model in the study by Van Cranenburgh and Garrido-Valenzuela (2025) has been trained based on residential location choice data –which involved street-level images– that model can be used to predict the utility derived from street-level conditions based on street-level images. Hence, it can be used to shed light on the spatial distribution of street-level conditions within a city.

The aims of this research are twofold: first, to shed light on the distribution of utility derived from street-level conditions in a residential location choice context at a city-level scale, and second, to examine the factors shaping this distribution. We use the city of Rotterdam as our case study. Rotterdam is the second largest city in the Netherlands, with about 670k inhabitants, and boasts a diverse array of neighbourhoods (Custers & Willems, 2024). To achieve our research aims, we capitalise on the widespread availability of street-level imagery and the recently developed CV-DCM. Specifically, we collect 300k geo-tagged street-level images of Rotterdam and calculate the utility derived from the street-level conditions using the CV-DCM. Finally, we aggregate the results at the postal code level to produce maps showing the spatial distribution.

This research makes a substantive and methodological contribution. Substantively, it is the first to present a large-scale application of a CV-DCM, demonstrating its potential to generate new insights about *preferences* of urban environments. As a methodological contribution, we extend CV-DCMs to deliver insights into the factors underlying the utility distribution derived from street-level conditions. Specifically, we have added a semantic regularisation layer to the model. This layer is designed to predict key semantic attributes that are believed to influence residential location choices, such as the number of cars and the amount of visible sky, alongside the location decisions themselves. Thus, the extended model extracts the semantic attributes from images, which it, in turn, uses to explain the residential location choices. This integrated approach eliminates the need for a separate pipeline to process, segment, or otherwise extract information from images, thereby streamlining the analysis. Moreover, because the extended model is computationally efficient to deploy, it enables the analysis of a large number of

images, making it feasible to assess the utility distribution derived from street-level conditions on a city-wide scale.

2. Method

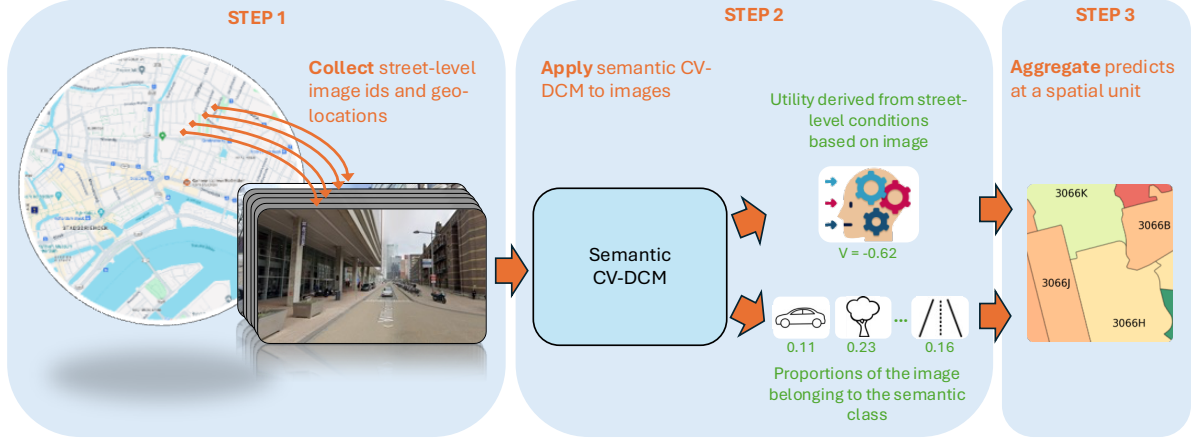Figure 1 shows an overview of our methodology, which comprises three main steps:

Figure 1: Overview of methodology

2.1. Step 1: Image data retrieval for Rotterdam

Step 1 of the method involves the collection of URLs to street-level images of residential areas in the city of Rotterdam. We use a similar approach as in (Van Cranenburgh & Garrido-Valenzuela, 2025). The final database contains of URLs from ~300k street-level images of residential streets in Rotterdam.

2.2. Step 2: Semantic CV-DCM

Step 2 involves applying a trained CV-DCM to the images collected in Step 1. To this end, this subsection introduces the CV-DCM (section 2.2.1), extends it (section 2.2.2), and reports on the training data (section 2.2.3) and training results (section 2.2.4) of the model that we apply.

2.2.1. CV-DCM

In this model, the utility of alternative *i*, denoted U_i, is derived from the numeric attributes X_i and attributes encoded in the image $\mathcal{I}_i$, which is presented as part of the alternative, see Figure 2. In line with the vast majority of discrete choice models, we assume utility is linear and additive, see Equation 1. The main advantage of these assumptions is that the β_m parameters can be interpreted as marginal utilities; representing the *ceteris paribus* change in utility associated with a one-unit change in the corresponding explanatory variable x_m. The mapping $g(\circ)$ can be further decomposed into two functions: first, $\varphi(\circ)$ produces a feature map (aka embedding) of the image, denoted $Z_{in} = \{z_{i1n}, z_{i2n}, \ldots, z_{iKn}\}$. This mapping is carried out by a computer vision architecture, such as a Convolutional Neural Network (CNN) or a Vision Transformer (ViT). After that, the feature map is mapped linearly onto utility: $\sum_k \beta_k z_{ikn}$.

3

$$U_{in}(X_{in}, \mathcal{I}_{in}) = \underbrace{\overbrace{\sum_m \beta_m x_{inm}}^{\substack{\textit{utility derived from} \\ \textit{numeric attributes}}} + \overbrace{\sum_k \beta_k z_{ikn}}^{\substack{\textit{utility derived from attributes} \\ \textit{encoded in image}}}}_{V_{in}} + \varepsilon_{in}$$

<div align="right">Equation 1</div>

$$\textit{where}$$
$$Z_{in} = \varphi(\mathcal{I}_{in})$$
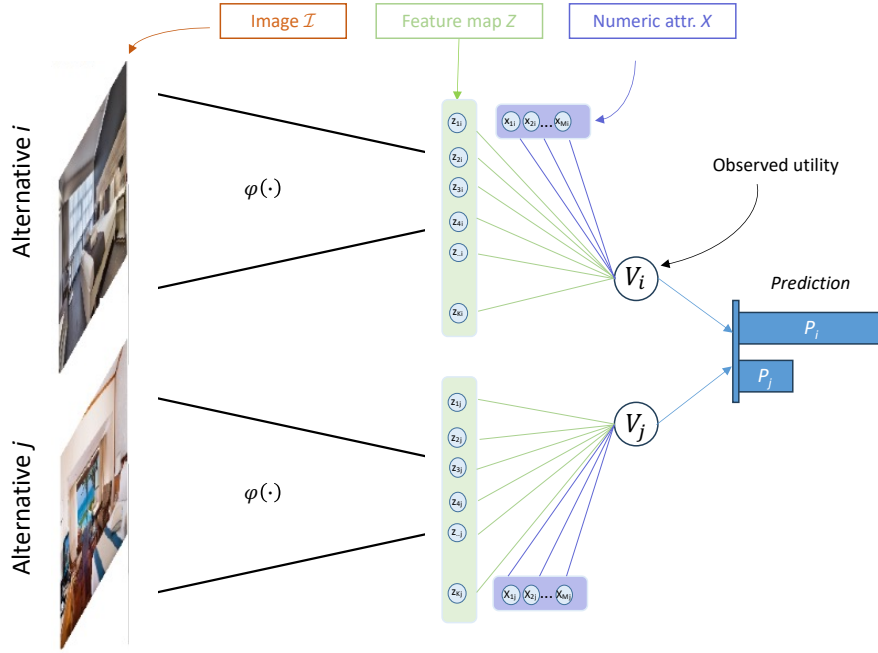$$\varepsilon_{in} = \sim EV\ type\ I$$

Figure 2: Model structure of CV-DCM (for J =2).
Subscript n dropped for legibility.

A drawback of the CV-DCM is that its computer vision part is opaque. Even though the feature map is mapped linearly onto utility (see Equation 2), the β_k parameters do not carry a behavioural meaning. This lack of interpretation is because the elements of the feature map, z_{ikn}, themselves do not carry a priori semantic meaning. Therefore, the CV-DCM does not allow tracing back which attributes encoded in a street-level image contribute to utility and to what extent.

One way to obtain more behavioural insights about people's preferences over the attributes encoded in street-level images (i.e. image feature map) is to extract semantic objects, like cars and trees, from the image and feed these directly into the utility function of a traditional choice model. Such an approach is taken in adjacent studies, like Nagata et al. (2020); Ramírez et al. (2021). This approach allows for estimating the effect of the attributes on the utility function, offering a clear understanding of how specific objects in the environment influence choices. Such a top-down approach, however, is constrained by the analyst's predefined selection of objects, limiting the ability to detect unexpected or emergent elements. Furthermore, it cannot capture more nuanced visual cues, such as Gestalt principles—proximity, similarity, common

fate, common region, closure, continuity, connectedness, and common orientation. These factors play a significant role in shaping how people perceive and interact with their surroundings and are widely applied in urban planning (Gjerde, 2011).

2.2.2. Semantic CV-DCM

In light of the considerations in section 2.2.1, we propose to extend the CV-DCM to provide more behaviour insights while keeping the model capable of capturing nuanced visual cues. Specifically, we propose adding a semantic regularisation layer to the model that aims to explicitly predict semantic attributes that are believed to be important to the residential location choice, such as the cars, sky and trees, see Figure 3. The semantic attributes (nodes) are predicted from the feature map, which, in turn, are assumed to map linear onto utility. Because the regularised nodes carry semantic meaning and are mapped linearly onto utility, the associated β^{sem} parameters can be interpreted as the marginal utilities of the semantic objects. Thus, by adding the semantic regularisation layer, the model captures the utility of the semantic attributes in an explicit and transparent way (as opposed to via β_k). Equation 2 formally introduces the Semantic CV-DCM.

$$U_{in} = \overbrace{\underbrace{\sum_m \beta_m^{num} x_{imn}}_{\substack{Utility\ derived\ from \\ numeric \\ attributes}} + \underbrace{\sum_t \beta_t^{sem} s_{itn}}_{\substack{Utility\ derived\ from \\ semantic\ attributes \\ encoded\ in\ the\ image}} + \underbrace{\sum_k \beta_k^{res} z_{ikn}}_{\substack{Residual\ utility\ derived \\ from\ attributes \\ encoded\ in\ the\ image}}}^{V_{in}} + \varepsilon_{in} \qquad \text{Equation 2}$$

$$where\ Z_{in} = \varphi(\mathcal{I}_{in}|w_\varphi)$$
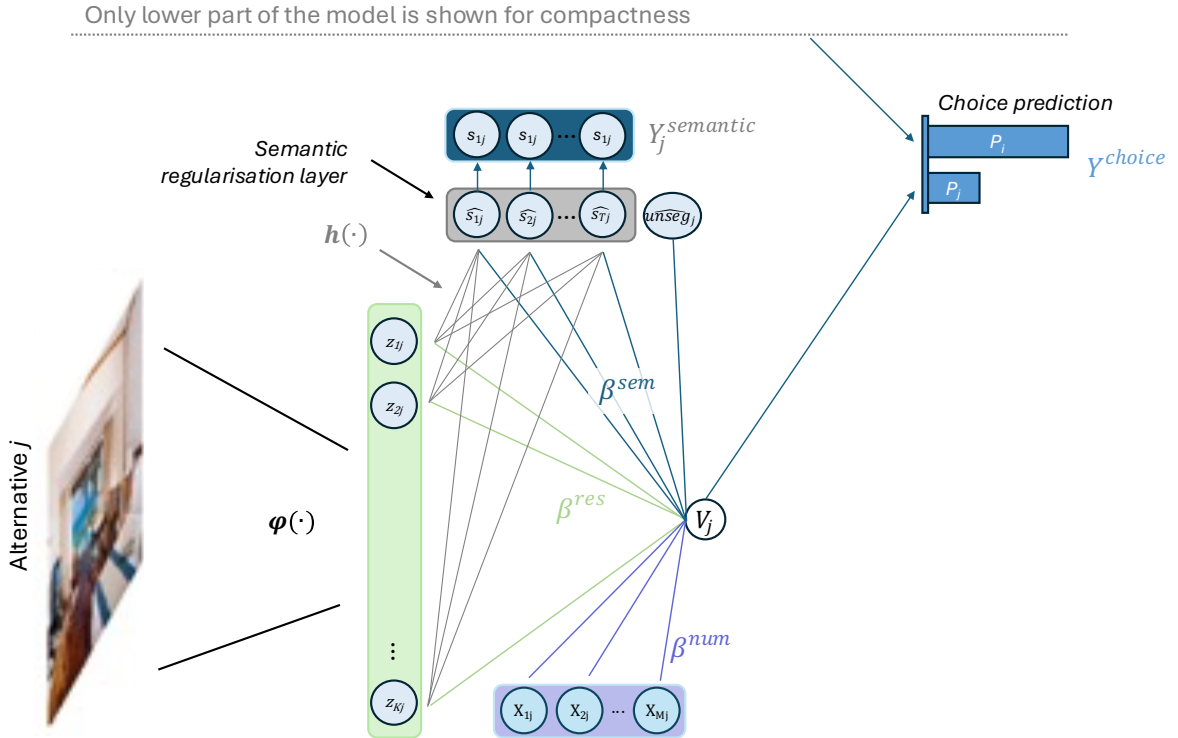
$$S_{in} = h(Z_{in}|w_h)$$

Figure 3: Semantic CV-DCM (for J =2).

5

2.2.3. Training data

To train the Semantic CV-DCM, we need two sources of data: (1) a data set containing the choices over choice tasks with numeric and visually encoded attributes, and (2) a data set containing ground truth levels of the semantic attributes encoded in the images used in the choice data set. Regarding the former, in this study, we use the residential location choice data collected by.[1] Figure 4 shows a screenshot of a choice task from their stated choice experiment.

Figure 4: Screenshot of the stated choice experiment (image source: Google)
(translated to English)

Regarding the data set containing the ground truth levels of the semantic attributes encoded in the images, we used state-of-the-art zero-shot object detection and segmentation models to construct it. Unlike traditional object detection models that have been trained using supervised learning, zero-shot models can detect objects beyond those predefined classes for which the models have been trained. More specifically, our detection and segmentation pipeline combines an object detection model called GroundingDINO and a segmentation model called Segment Anything Model (SAM). GroundingDINO is a zero-shot model for detecting objects developed by Liu et al. (2023). It integrates a transformer-based detection model called DINO (Zhang et al., 2022) and grounded pre-training. DINO achieves object detection based on text prompts. We applied this pipeline to all images used in the stated choice experiment, using a list of semantic attributes informed by a brief literature review. Figure 5 illustrates four examples of the segmented images based on this list of segments.

---

[1] https://github.com/TUD-CityAI-Lab/Computer-vision-enriched-DCMs

Figure 5: Four examples of segmented images using GroundingDINO and SAM

2.2.4. Training results

Table 2 shows the training results related to the choice data. The first column reports the benchmark model, i.e. the plain-vanilla CV-DCM of Van Cranenburgh and Garrido-Valenzuela (2025); the second column shows the results of the Semantic CV-DCM. Looking at the model fit in the test data set, we see that the Semantic CV-DCM fits the data almost equally well as the benchmark model. The fact that the benchmark CV-DCM outperforms the Semantic CV-DCM is expected because the CV-DCM imposes less structure on how the utility is derived from attributes encoded in the image. For instance, the benchmark CV-DCM may learn a nonlinear relation between the number of cars and utility. In contrast, the Semantic DCM imposes a linear relationship. The comparatively better fit on the training dataset suggests that the model slightly overfitting to the training data.

Next, we look at the interpretable parameters of the Semantic CV-DCM. Firstly, in line with expectations, we see that the estimates associated with the numeric attributes – housing cost and commute travel time – are virtually the same as those of the benchmark model. Secondly, the signs of the semantic parameters have the intuitively expected signs. In this application, we fixed $\beta_{building}^{sem}$ to zero for normalisation. Given this normalisation choice, we expect positive signs for the parameters associated with e.g. trees, grass, and sky. After all, a positive parameter for a semantic attribute means that increasing its proportion – at the expense of the proportion of buildings (which is the reference category) – will increase the utility derived from the street-level conditions. Likewise, we expect negative signs for the parameters associated with, e.g. roads and fences, because increasing their proportion – at the expense of buildings – can be expected to decrease the utility derived from the street-level conditions.

7

Table 2: Training results semantic CV-DCM

Data set			CV- DCM (benchmark)	Semantic CV-DCM
Train data N = 9,784		Parameters	86m	86m
		Log-likelihood	-5,724	-5,572
		ρ^2	0.156	0.17
		Cross entropy	0.585	0.570
		Computation time	~1.5 hr	~6.0 hr
Test data N = 1,948		Log-likelihood	-1137.6	-1145.4
		ρ^2	0.158	0.152
		Cross entropy	0.585	0.588
Interpretable model parameters	Numeric Attr.		Est	Est
		β_{hhcost}	-0.94	-0.94
		β_{tt}	-0.23	-0.24
	Semantic Attr.	β_{car_c} (count)		-0.25
		β_{car_p} (proportion)		-0.59
		$\beta_{building}$ (proportion)		0.00*
		β_{grass} (proportion)		0.96
		β_{road} (proportion)		-0.59
		β_{sky} (proportion)		1.42
		β_{trees} (proportion)		1.40
		β_{plants} (proportion)		1.05
		B_{fence} (proportion)		-0.81
		B_{water} (proportion)		0.13
		$B_{unsegmented}$ (proportion)		-0.25

* *Fixed to zero for normalisation*

2.3. Step 3: Model application and aggregation

In the application, we utilise the trained Semantic CV-DCM model to compute utilities derived from the street-level conditions based on the street-level images of Rotterdam. Notably, in the application, we exclude utilities associated with the numeric attributes 'commute travel time' and 'housing cost', which were part of the training data but are unimportant to our application, in which we want to examine the spatial distribution of utility derived from street-level conditions. Excluding the numeric attributes thus allows us to isolate the utility levels that reflect the street-level conditions.

3. Results: spatial distribution of residential street-level condition

The middle plot of Figure 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 low utility derived from the residential street-level conditions, and the colour green indicates a high utility derived from the residential street-

level conditions. Figure 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 environment, 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, respectively, the presence of the museum district and a medium-sized city park.

Secondly, street-level conditions vary considerably, even within small areas. For example, in the neighbourhood 'Hillegersberg Noord' (at the top of the plot), a single dark red spot stands out in an otherwise attractive area, illustrating how conditions can rapidly improve or deteriorate even within close proximity. So, even though 'Hillegersberg Noord' is generally considered an upscale and attractive neighbourhood, this area contains pockets where the street-level conditions are not appealing.

Thirdly, 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.

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.
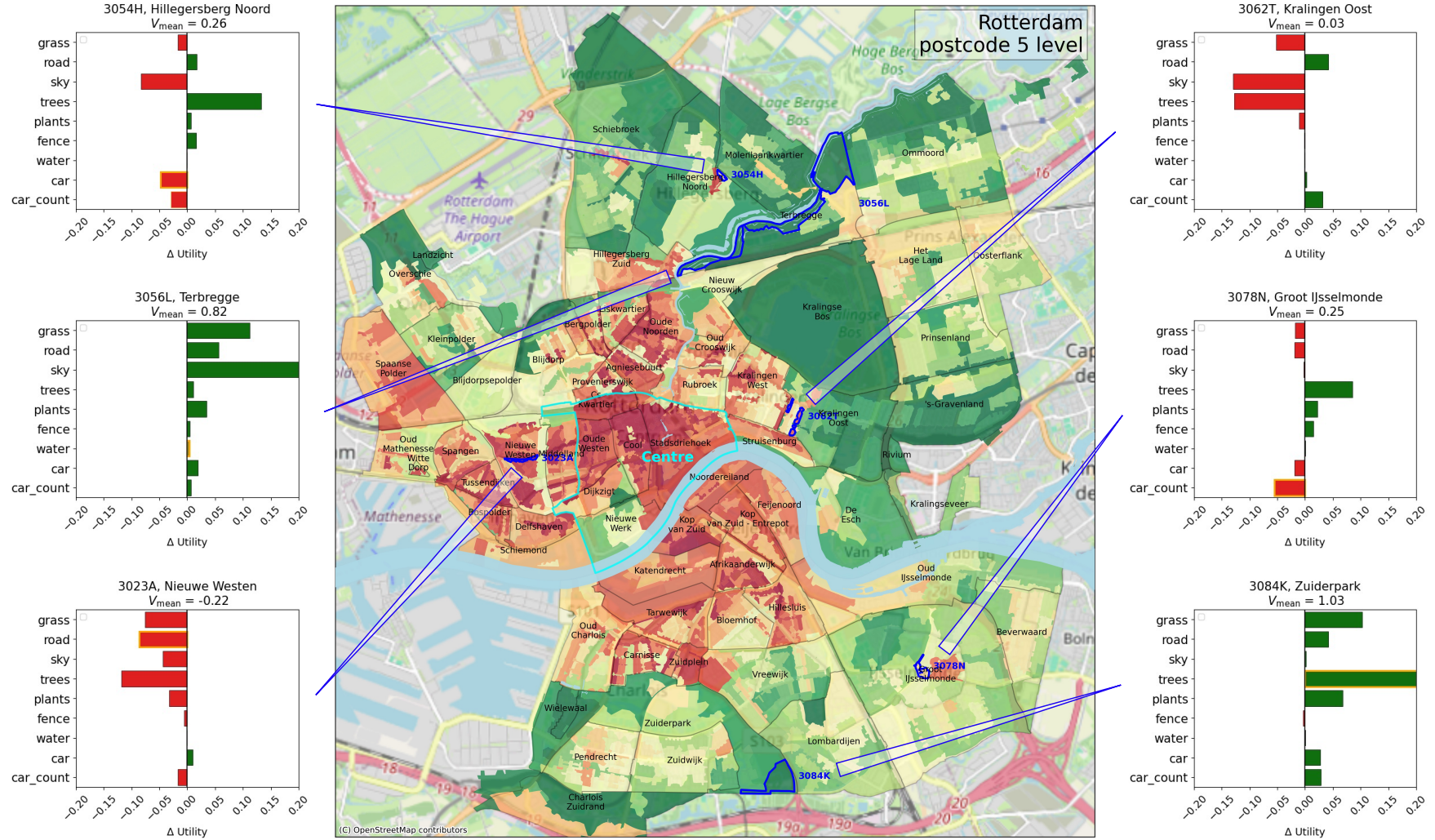
Figure 7: Spatial distribution of residential street-level conditions utility (from images) in Rotterdam (middle). Locations of areas with the highest proportion of semantic attributes are encircled in blue. Impact on street-level utility compared to the mean is shown on both sides for selected areas.

10

4. Conclusion and discussion

This paper utilises recently proposed computer-vision enriched discrete choice models (CV-DCMs) to provide novel insights into the spatial distribution of utility derived from street-level conditions. Thereby, it advances the residential location choice literature by offering insights into the role street-level conditions play in these decisions. Additionally, it complements previous research on urban environment perceptions based on street-level imagery by introducing a preferences-based counterpart. Finally, it makes a methodological contribution by extending the CV-DCM models with a semantic regularisation layer. This layer extracts key semantic attributes from images, which the model then uses to predict the residential location choices. Thereby, we elucidate the computer vision part of the CV-DCM and eliminate the need for separate image processing pipelines. The next steps involve further analysis of the results and a deeper examination of the proposed method to enhance its ability to understand residential location choices.

References

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

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*, 105009.

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, *Computer Vision – ECCV 2016* Cham.

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.

Gjerde, M. (2011). Visual evaluation of urban streetscapes: How do public preferences reconcile with those held by experts? *Urban design international*, *16*, 153-161.

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

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

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems*, *65*, 113-125.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Li, C., Yang, J., Su, H., & Zhu, J. (2023). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. *arXiv preprint arXiv:2303.05499*.

Ma, X., Ma, C., Wu, C., Xi, Y., Yang, R., Peng, N., Zhang, C., & Ren, F. (2021). Measuring human perceptions of streetscapes to better inform urban renewal: A perspective of scene semantic parsing. *Cities*, *110*, 103086.

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

Nagata, S., Nakaya, T., Hanibuchi, T., Amagasa, S., Kikuchi, H., & Inoue, S. (2020). Objective scoring of streetscape walkability related to leisure walking: Statistical modeling approach with semantic segmentation of Google Street View images. *Health & place*, *66*, 102428. https://doi.org/https://doi.org/10.1016/j.healthplace.2020.102428

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.

Ramírez, T., Hurtubia, R., Lobel, H., & Rossetti, T. (2021). Measuring heterogeneous perception of urban space with massive data and machine learning: An application to safety. *Landscape and Urban Planning*, *208*, 104002.

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.

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*.

Wei, J., Yue, W., Li, M., & Gao, J. (2022). Mapping human perception of urban landscape from street-view images: A deep-learning approach. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102886.

Zhang, A., Song, L., & Zhang, F. (2022). Perception of pleasure in the urban running environment with street view images and running routes. *Journal of Geographical Sciences*, *32*(12), 2624-2640.

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.