Using computer vision-enriched discrete choice models to assess the visual impact of transport infrastructure renewal projects:
A case study of the Delft railway zone

Sander van Cranenburgh[1]
Francisco Garrido Valenzuela[1]
[1]CityAI lab, Transport and Logistics Group, Delft University of Technology

Abstract
a computer vision-enriched discrete choice model to investigate the impact of the redevelopment of the Delft railway zone on the visual environment. Using computer vision-enriched discrete choice models, we evaluate the changes in the utility levels derived from the visual environment by analysing over 70k street-view images from periods before and after the redevelopment of the railway zone in Delft. We find evidence that the visual appearance of the railway zone has considerably improved after the redevelopment project. This finding highlights the potential of using computer vision-enriched discrete choice models to quantitatively evaluate and monitor changes to the visual environment arising from new transport infrastructure projects.

Keywords
Discrete choice modelling, Appraisal, Computer vision, Street-view images

1. Introduction

Cost-Benefit Analysis (CBA) involves tallying up all costs of a (transport) project and subtracting that amount from the total projected benefits of the project. Since the benefits are often not expressed in euros, the latter requires a monetisation step to convert these to euros. Some benefits of transport projects are comparatively easy to monetise, such as travel time savings. Over the years, an extensive practice has been established (Small 2012; Kouwenhoven et al. 2014). Other effects are still more challenging to monetise, often because of their abstract or enigmatic nature. One key example of a hard-to-monetise benefit involves changes to the visual appearance and environment. Transport projects often have a major (visual) impact on the landscape. Their visual impact often plays a crucial role in the political debate leading to the decision to build new transport infrastructure.

However, the visual impacts of a transport project are typically merely assessed qualitatively. As a result, they are not included in either of the indicators of CBA that are decisive in the political process: the benefit-cost ratio or the net present value (Annema and Koopmans 2015). This weak position of impacts on the visual environment and landscapes, more generally, can lead to poor land-use decisions that cause welfare losses which undermine public support.

Recently, Computer Vision-enriched Discrete Choice Models (henceforth: CV-enriched DCMs) have been proposed (Van Cranenburgh and Garrido-Valenzuela 2023). This model extends the application of discrete choice models towards visual preferences. Van Cranenburgh and Garrido-Valenzuela (2023) demonstrate their use in residential location choices – showing how trade-offs are captured between monthly cost, travel time and street-level factors, such as openness, building typology and greenness (as embedded in images).

This study applies the CV-enriched DCM trained by (Van Cranenburgh and Garrido-Valenzuela 2023) to investigate changes to the visual environments resulting from transport infrastructure projects.

Specifically, we focus on the Delft railway zone, which was put underground in the period 2014-2015. This transformed the visual appearance of the whole railway zone. Using the trained CV-enriched DCM, we compute utility levels for over 70k Google street-view images from before and after the redevelopment of the railway zone. Thereby, we aim to provide a rigorous quantitative underpinning of the benefits to the visual environment arising from the redevelopment of the railway zone. With this work, we contribute to the stream of research that capitalises on street-view images as a source of information about the urban environment (Naik et al. 2017; Rossetti et al. 2019; Ma et al. 2021; Ramírez et al. 2021; Garrido-Valenzuela et al. 2022)

2. Methodology

Our method involves the following four steps. First, we collect street-view images before and after the infrastructure renewal in the surroundings of Delft railway station. Second, we apply the CV-enriched DCM trained by Van Cranenburgh and Garrido-Valenzuela to the images to produce a utility level per image. Third, we aggregate the utilities across spatial hexagons for spatial analysis. Fourth, we analyse the changes in aggregate utility levels before and after the infrastructure renewal.

2.1. Delft railway station area and image data collection

After years of fierce political debate, in the early 2000s, the final decision was made to redevelop the railway zone in Delft. A significant part of the project involved putting 2.3km of railway track and the train station underground. The main construction period on the railway track took place in 2014 and 2015. Figure 1 shows the railway station before and after the redevelopment. Much of the debate leading up to the commissioning of the project was about whether the improvement in the visual environment in the railway zone actually would exceed the considerable construction costs. Because of the significant financial burden the project turned out to be, even today, the project's benefits are debated.

Figure 1: Delft Railway station before (left) and after (right) redevelopment

To collect street-view images in the Delft railway zone, we created a grid of points with 25-metre spacing in an 800m circumference around the city centre. We retrieved the nearest street-view image id for each point on the grid using Google's street-view API. If multiple years were available, we collected images of all available years. Each street-view image id corresponds to a 360-degree panorama view at the street level. Finally, from each panorama, we generated two image urls with 90-degree angles to the direction of the street. This latter ensures the images are side-views (e.g., as opposed to views parallel to the driving direction of the Google car). All street-view images are (temporally) stored using png format with 900 x 600 pixels and 8 bits per colour channel (implying 16.7m colour values per pixel). For each image, the geo-location, year and month are stored. Figure 2 shows the collected number of street-view urls per year.

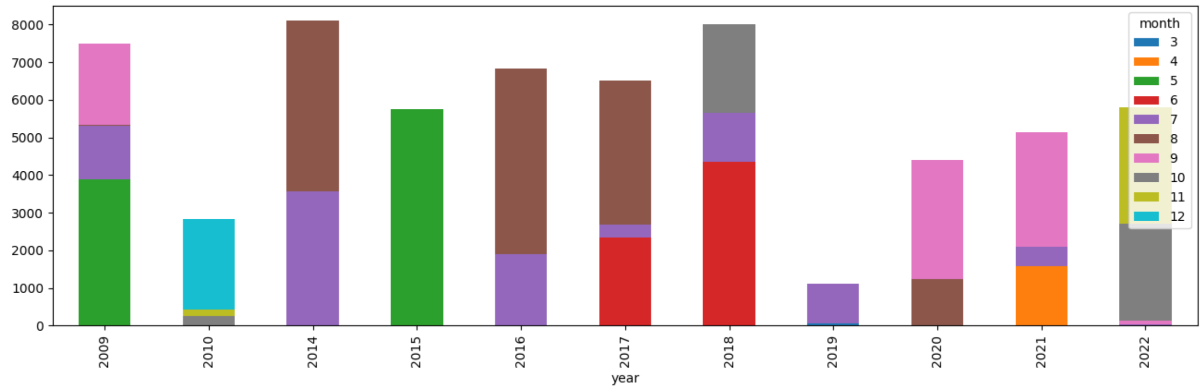

Figure 2: Number of street-view images per year

2.2. Computer vision-enriched discrete choice models

To obtain the utility levels of each image, we apply the CV-enriched DCM trained by Van Cranenburgh and Garrido-Valenzuela (2023) on residential location choice data. This model assumes decision-makers, denoted *n*, make decisions based on Random Utility Maximising (RUM) principles (McFadden 1974). Equation 1 shows the utility function of this model we use. As can be seen, in this model, the utility, U_j, is derived from the numeric attributes X_j and attributes embedded in the street-view images S_j, which were presented as part of the alternatives. Furthermore, the account for the possibility that images taken, e.g. in spring look, on average, more attractive than images taken in winter, a constant per month is included in the utility function (the second term from the left), where I_{Sj} is a binary vector with value one if the image is taken in month *mo*, and zero otherwise.

Figure 3 shows a screenshot of the stated choice data on which the CV-enriched DCM is trained. In this experiment, respondents had to make trade-offs between street views (visual appearance of the neighbourhood) and two numeric attributes: monthly housing cost and commute travel time. The street-view images shown to respondents in choice tasks were randomly drawn from an extensive database of street-view image ids. Based on these data, the CV-enriched DCM could learn the preferences over elements embedded in the street-view images, such as compactness, openness, street topology, parking facilities, etc. For more details about the data collection and the CV-enriched DCM, see Van Cranenburgh and Garrido-Valenzuela (2023).

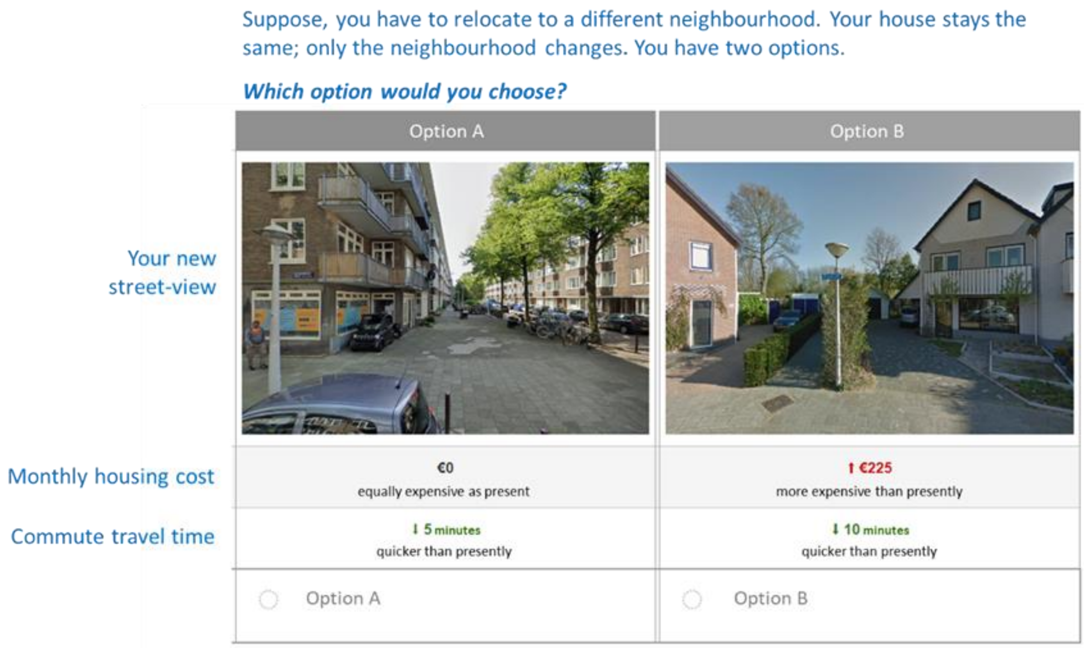

Figure 3: Screenshot of stated choice experiment used to train the CV-enriched DCM

Importantly, in Equation 1, φ is a mapping function – performed by the computer vision part of the model: DeiT base (Touvron et al. 2021) – which maps image S_j onto a lower dimensional feature space, denoted Z_j (which has a dimensionality of 1 x K). The feature map of an image embeds the relevant information from that image that generates (dis)utility and, in turn, maps linearly map onto the utility. w_k denotes the weight associated with the k^{th} feature of Z_j; β_m denotes the marginal utility associated with attribute m, and x_{jmn} denotes the attribute level of numeric attribute m of alternative j, as faced by decision-maker n.

$$U_{jn} = \overbrace{\sum_m \beta_m x_{jmn}}^{\substack{\textit{Utility derived} \\ \textit{from numeric attributes}}} + \overbrace{\sum_{mo} \beta_{mo} I_{S_j}}^{\substack{\textit{Utility derived} \\ \textit{from the month of the year}}} + \overbrace{\sum_k w_k z_{jkn}}^{\substack{\textit{Utility derived} \\ \textit{from image feature map}}} + \varepsilon_{jn} \quad where\, Z_{jn} = \varphi\left(S_{jn} \mid w_r\right) \qquad \text{Equation 1}$$

In this application of the model, we are solely interested in the utility it produces from images. In other words, to deploy this model, we first apply the mapping function φ to images to obtain feature maps Z. In turn, we take the inner product with w to get the utility of the image. Finally, we 'correct' the utility of each image for the month of the year in which the image was taken (the second term of Equation 1). Figure 4 kernel density plots of the utility levels computed from the images (before and after the renewal project). The left-hand side plot shows utility levels uncorrected for the month of the year; the right-hand side plot shows utility levels corrected for the month of the year.

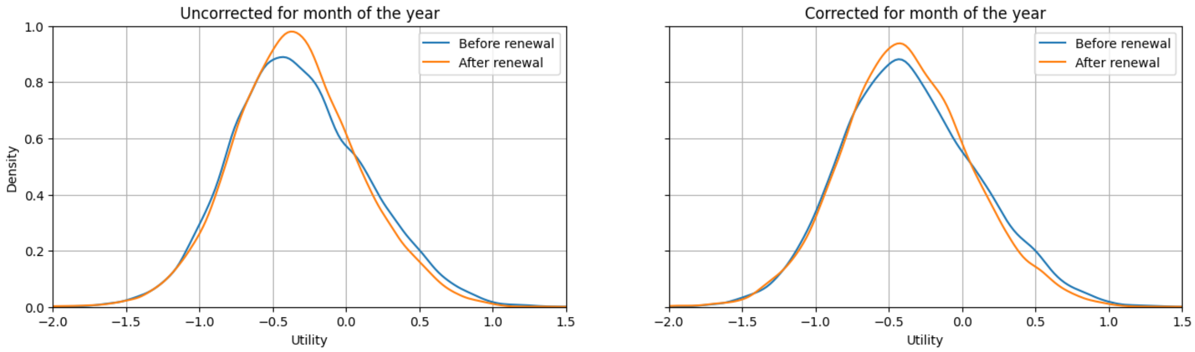

Figure 4: Kernel density plots of utility levels before and after the renewal project.

2.3. Aggregation

To investigate the potential changes in utility levels arising from the changes in the visual environment arising from the infrastructure project, we must define a spatial unit of analysis. For this purpose, we use regular hexagonal cells with 25-metre sides to tessellate the entire study area. Accordingly, on average, each hexagon contains 20 images from the period before and 15 images from the period after the redevelopment.

3. Preliminary results

Figure 5 shows the main results of this study. The hexagon's colour depicts how the utility level changed between the periods before and after the redevelopment of the railway zone. A green colour indicates a positive change in the average utility derived from the images within the hexagon; a red colour indicates a negative change in the average utility. The map shows an area of approximately 2 x 2 km. Delft has two landmark churches, which are depicted to ease navigation.

Based on Figure 5, we make a number of observations. Firstly, the visual appearance of the redeveloped area has considerably improved. Almost all hexagons within the red encircled area are greenish, implying a positive change in utility. This is in line with behavioural intuition. It also suggests that CV-enriched DCMs can indeed be used to evaluate changes in the visual environment. Secondly, the visual appearance of the area west of the redeveloped railway zone has also improved. Again, we see predominantly green hexagons. A possible explanation for this observation is that the redeveloped railway zone radiated positively in this direction and led to positive changes to the visual environment. Thirdly, the change in the visual appearance of the inner city (located Nord-East of the train station) is mixed. Some streets seem to have deteriorated (coloured orange and red), while others show positive changes in their visual appearance (coloured green). Presumably, local explanations can be found explaining these changes.

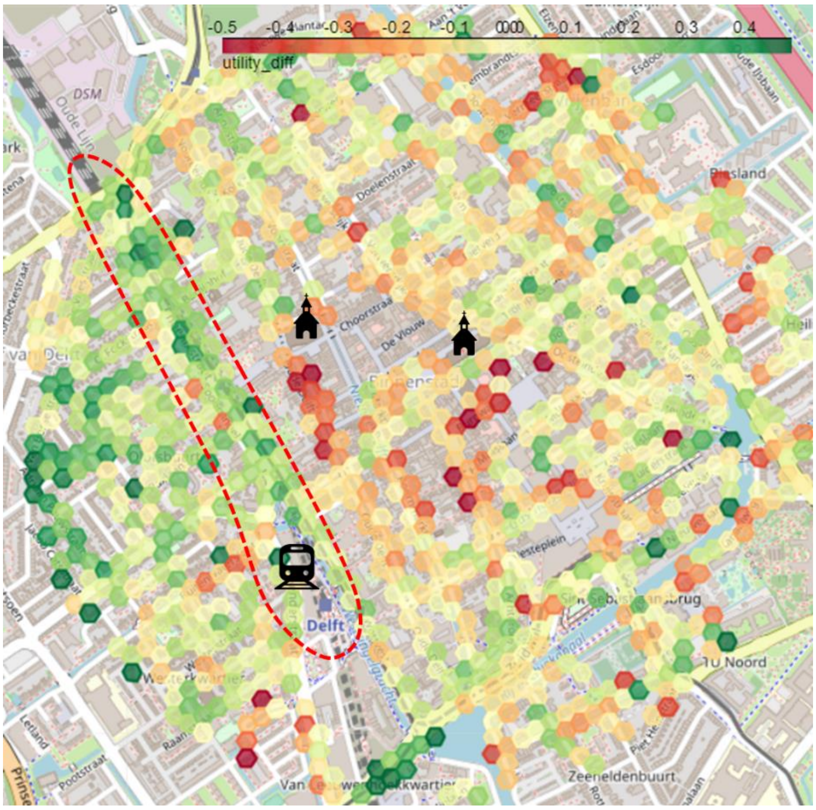

Figure 5: Changes in utility levels Delft. The railway zone is encircled in red.

4. Conclusion and discussion

This research has employed new computer vision-enriched discrete choice models to investigate changes in the visual environment arising from the redevelopment of the Delft railway station zone. The intuitively correct results we obtained from our case study suggest that CV-enriched DCMs can indeed be used to evaluate changes in the visual environment.

A key limitation is this research is that the utilities extracted from the CV-enriched DCM reflect the attractiveness as a residential location. But, the function of the railway station zone is mixed. Its functions include housing, transfer, gathering, working, eating, etc. In this research, we only looked at the visual environment from the lens of residential location.

Next steps

We envision taking the following steps in the coming months. Firstly, we would like to expand our study areas. We want to apply the approach to other areas that undergo renewal to establish its robustness and further applicability. In addition, we aim to develop a better grasp of optimal hexagon size in combination with data availability (Wong 2004). Finally, we aim to show what the trained CV-enriched DCM has learned. We want to understand what causes exceptionally high or low utility predictions. Such model explanations may help to better inform urban planners and policymakers on future transport infrastructure projects.

Acknowledgement

This work is supported by the TU Delft AI Labs programme.

References

Garrido-Valenzuela, F., van Cranenburgh, S. & Cats, O. (2022). Enriching geospatial data with computer vision to identify urban environment determinants of social interactions. *AGILE: GIScience Series*, *3*, 72.

Kouwenhoven, M., de Jong, G. C., Koster, P., van den Berg, V. A. C., Verhoef, E. T., Bates, J. & Warffemius, P. M. J. (2014). New values of time and reliability in passenger transport in The Netherlands. *Research in Transportation Economics*, *47*, 37-49.

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, D. (1974). *The measurement of urban travel demand*. (Berkeley: Institute of Urban & Regional Development, University of California).

Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L. & Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, *114(29)*, 7571-7576.

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.

Rossetti, T., Lobel, H., Rocco, V. & Hurtubia, R. (2019). Explaining subjective perceptions of public spaces as a function of the built environment: A massive data approach. *Landscape and Urban Planning*, *181*, 169-178.

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation*, *1(1)*, 2-14.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A. & Jégou, H. (2021). Training data-efficient image transformers & distillation through attention. International Conference on Machine Learning, PMLR.

Van Cranenburgh, S. & Garrido-Valenzuela, F. (2023) Computer vision-enriched discrete choice models to explore the importance of street-level factors and commute travel times in the residential location choice behaviour,

Wong, D. W. (2004). The modifiable areal unit problem (MAUP). *WorldMinds: geographical perspectives on 100 problems: commemorating the 100th anniversary of the association of American geographers 1904–2004*, 571-575.