



Modelling Preferences on Street Redesigns

A Perception-Driven Framework Using Visual and Segmentation Data

Miguel Gouveia Pinto Leite Alves

Thesis to obtain the Master of Science Degree in

Civil Engineering

Supervisors: Prof. Dr. Filipe Manuel Mercier Vilaça e Moura

Prof. Dr. Gabriel Costa Valença

Examination Committee

Chairperson: Prof. Jorge Manuel Lopes Baptista e Silva

Supervisor: Prof. Filipe Manuel Mercier Vilaça e Moura

Members of the Committee: Prof. Carlos Miguel Lima de Azevedo

October 2025

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa. Declaro ainda que foram utilizadas ferramentas de Inteligência Artificial, nomeadamente o ChatGPT-5 e o Grammarly, para melhorar a clareza do texto, apoiar na elaboração de scripts e gerar imagens base posteriormente modificadas manualmente. Estas ferramentas não contribuíram para a criação do conteúdo de investigação nem para as conclusões apresentadas. Assumo total responsabilidade por todo o material presente nesta dissertação.

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa. I further declare that Artificial Intelligence tools, namely ChatGPT-5 and Grammarly, were used to improve textual clarity, assist with scripts and generate base images later modified by hand. These tools did not contribute to the creation of research content or conclusions. I take full responsibility for all material presented in this thesis.

Acknowledgements

This thesis would not have been possible without the support, guidance and encouragement I received throughout its development and across my academic journey.

First and foremost, I would like to express my sincere gratitude to my supervisors. To Professor Gabriel Valença, for his practical guidance and proactive involvement at every stage of this work. To Professor Filipe Moura, for having accompanied me since my first steps in the field of transport engineering and for inspiring me to change my area of specialisation.

I am deeply thankful to the U-Shift Research Group for the inspiring environment and constant intellectual exchange that defined this experience from the very beginning. I am equally grateful to the Foundation for Science and Technology (FCT), the University of Coimbra and Instituto Superior Técnico, which funded the project “*STREETS4ALL – Dynamic and equitable Reallocation of Urban Street Space*”, making this work possible and fostering innovation in urban mobility research. To all my colleagues and friends at U-Shift, thank you for creating such a welcoming and stimulating atmosphere.

I would also like to thank all the friends I met throughout my academic path, whose companionship made even the most demanding moments enjoyable. A special thanks to Pedro, Diogo and both Franciscos, who have been by my side since the very beginning and with whom I shared countless projects, challenges and laughs throughout both degrees. To my high-school friends who came along for this journey, Penedos, Guilherme, Bonnet, Machado, Costa and Tiago, thank you for your continued presence and friendship through all these years.

My deepest gratitude goes to my family. To my mom and dad, who have given me every condition to pursue what I wished, always finding the perfect balance between freedom and support. To my sister, for your help and encouragement throughout this journey. And finally, to Madalena, for being my greatest motivation and my constant source of calm through all the highs and lows.

Abstract

Urban streets are increasingly redesigned to prioritise sustainable mobility, yet planning processes often overlook how citizens visually perceive and emotionally interpret these transformations. This gap can generate resistance to otherwise beneficial interventions. To address this challenge, this dissertation developed a perception-based analytical framework linking measurable visual changes in street environments to subjective evaluations. Using temporal images of real interventions, semantic segmentation and perceptual colour metrics were applied to quantify spatial and chromatic variations such as vegetation gain, surface redistribution and hue balance. These indicators were integrated with survey-derived perceptual ratings through factor analysis and structural equation modelling to identify the latent constructs and to predict preference of redesigned streets.

Results showed that both colour and segmentation-based predictors captured distinct yet complementary perceptual mechanisms: colour diversity and brightness enhanced liveability and comfort, while balanced functional composition supported perceptions of safety and order. The framework yielded predictions across demographic and behavioural groups, revealing exploratory moderation effects and offering interpretable guidance for design. Despite limitations related to sample scope, illumination variability and segmentation bias, the approach demonstrated the feasibility of operationalising perception within urban analytics.

Overall, the study advanced the integration of visual data and human perception into evidence-based planning, providing a replicable tool to anticipate public reactions and guide participatory, sustainable urban transitions.

Keywords: urban perception; street design; perceptual colour metrics; sustainable mobility; participatory planning; exploratory moderation terms.

Resumo

As ruas urbanas estão a ser progressivamente redesenhadas, priorizando a mobilidade sustentável, mas os processos de planeamentos negligenciam, muitas vezes, a forma como os cidadãos as avaliam visualmente e interpretam emocionalmente. Esta lacuna pode gerar resistência a intervenções que, de outro modo, seriam benéficas. Respondendo ao dilema, esta dissertação desenvolveu um processo baseado na percepção, relacionando alterações visuais mensuráveis nos espaços urbanos com avaliações subjetivas. Utilizando imagens temporais de intervenções reais, aplicaram-se técnicas de segmentação semântica e métricas de cor perceptiva, quantificando variações espaciais e cromáticas, como o aumento de vegetação, a redistribuição de superfícies e o equilíbrio tonal. Estes indicadores foram integrados com avaliações perceptivas obtidas por inquéritos, recorrendo à análise fatorial e à modelação de equações estruturais, identificando termos latentes e prevendo a preferência por ruas redesenhadas.

Os resultados demonstraram que os preditores baseados em cor e segmentação captaram mecanismos perceptivos distintos, mas complementares: a diversidade e o brilho aumentaram a habitabilidade e o conforto, enquanto uma composição funcional reforçou as percepções de segurança e ordem. A metodologia produziu previsões entre grupos demográficos e comportamentais, revelando efeitos de moderação exploratórios e oferecendo orientações interpretáveis para o desenho urbano. Apesar das limitações associadas à amostra, à variabilidade de iluminação e à parcialidade da segmentação, a abordagem demonstrou a viabilidade de operacionalizar a percepção na análise urbana.

No global, o estudo promoveu a integração de dados visuais e da percepção humana no planeamento urbano, oferecendo uma ferramenta replicável para antecipar reações públicas e orientar transições urbanas sustentáveis e participativas.

Palavras-Chave: percepção urbana; desenho urbano; métricas de cor perceptiva; mobilidade sustentável; planeamento participativo; termos de moderação exploratória.

Table of Contents

Acknowledgements	IV
Abstract.....	VI
Resumo	VIII
List of Figures	XII
List of Tables	XIV
List of Equations	XVI
List of abbreviations and definitions	XVIII
1 Introduction	1
1.1 Motivation.....	1
1.2 Problem definition and research questions.....	2
1.3 Dissertation structure	3
2 State of the art	5
2.1 Perceptions and public preference on street design	5
2.2 Image-based urban sensing and participatory evaluation methods	6
2.3 Colour and visual metrics used in perception research.....	7
2.4 Computer vision for street-scene analysis.....	9
2.5 Methodological concerns for before and after image comparison.....	10
2.6 Psychometrics and statistical modelling	11
2.7 Summary and Research Gaps	12
3 Methodology	15
3.1 Methodological framework.....	15
3.2 Survey data.....	17
3.3 Image collection and preprocessing	18
3.4 Colour data gathering	20
3.5 Segmentation model development	24
3.6 Segmentation inference and variation measurement.....	31
3.7 Preliminary statistical analysis	33
3.7.1 Univariate analysis	33
3.7.2 Bivariate analysis	35

3.7.3 Polychoric factor analysis on perception variables	37
3.8 Structural equation modelling	40
3.8.1 Base models.....	41
3.8.2 Moderation models.....	43
3.9 Predictive tool	46
3.10 Demonstration of applied value	50
4 Results and discussion.....	53
4.1 Image analysis	53
4.2 Survey responses	56
4.3 Translating visual data into perceptions	57
5 Limitations and future work.....	65
5.1 Implications and limitations	65
5.2 Future directions	67
5.3 Final remarks	68
References	71
Appendices	77
Appendix A	77
Bivariate analysis: perceptions x colour and segmentation variables (Spearman's rho)	77
Appendix B	78
Bivariate analysis: correlation between all colour and segmentation variables. (Pearson's r).....	78
Appendix C	79
SEM effects directional ranking per latent construct (Functionality, colour system).	79

List of Figures

Figure 1 - Illustrated summary of the methodology.	16
Figure 2 - Example before and after image pair shown to respondents. (Inês Duarte Lopes, 2025).....	17
Figure 3 – Likert scale evaluation of influence of the seven emotions. (Inês Duarte Lopes, 2025).....	17
Figure 4 – Google Street Views copyright restrictions. (Google, 2025)	18
Figure 5 – Graphical summary of the image sources used per intervention.	19
Figure 6 – Comparison between standard lab and empirically chosen thresholds.	21
Figure 7 – Representation of the colour presence bins and thresholds applied.....	22
Figure 8 – Shannon entropy: interpretability example.	22
Figure 9 – Intervention pair 101 and respective brightness data summary plot.	23
Figure 10 – Intervention pair 101 and respective saturation data summary plot.	23
Figure 11 – Representation of a pixel-wise segmentation step.....	24
Figure 12 – Representation of the data augmentation steps used.	26
Figure 13 – Segmentation model training evolution.....	28
Figure 14 - Map of leftImg8bit locations from the Cityscapes Dataset.(Cityscapes Dataset, 2025).....	29
Figure 15 – Intersection over Union representation.....	30
Figure 16 - Semantic segmentation model’s setup comparison.	31
Figure 17 – Semantic segmentation’s correction stage example.	32
Figure 18 – Example of a percentage of pixel-class change plot.....	33
Figure 19 - Example of univariate analysis summary for continuous variables.	34
Figure 20 - Univariate analysis of categorical variables (e.g. main commute mode).	35
Figure 21 - Polychoric correlation matrix for perception variables.	36
Figure 22 - Mosaic plot between x (safety) and y (urban specialist) variables.	37
Figure 23 - Parallel analysis’ scree plot.	38
Figure 24 - Representation of the EFA procedure on perception variables.	39
Figure 25 – Latent construct scheme based on the EFA and CFA processes.	40
Figure 26 - Representation of the PCA on brightness and saturation variables.....	41
Figure 27 - Representation of the segmentation-derived base SEM model.	43
Figure 28 - Representation of the residual-centred interaction term process.	44
Figure 29 - Simplification of the moderation SEM for colour or segmentation predictors.	45
Figure 30 - Representation of the latent construct prediction process.	49
Figure 31 - Alternative interventions to Avenida Rio de Janeiro, Lisbon.	51
Figure 32 - Barcelona superblock intervention example.....	54
Figure 33 - Survey respondents’ nationality reference.	56
Figure 34 - SEM effects directional ranking (Preference on segmentation and colour systems).	58

List of Tables

<i>Table 1 – Survey variables used and respective variable types.....</i>	<i>18</i>
<i>Table 2 – Colour variables used and respective variable types.</i>	<i>23</i>
<i>Table 3 – Shrink factor tuning results for the colour-related workflow.....</i>	<i>48</i>
<i>Table 4 – Colour and Segmentation overall preference predictions (Sample size: 1191).</i>	<i>59</i>
<i>Table 5 – Preference prediction by age group in the colour system.</i>	<i>60</i>
<i>Table 6 – Preference prediction by age group in the segmentation system.</i>	<i>60</i>
<i>Table 7 – Preference prediction by gender in the colour system.....</i>	<i>61</i>
<i>Table 8 – Preference prediction by place of residence in the colour system.</i>	<i>61</i>
<i>Table 9 – Preference prediction by main commute mode in the segmentation system.</i>	<i>62</i>
<i>Table 10 – Preference prediction by family composition in the segmentation system.....</i>	<i>62</i>
<i>Table 11 – Preference prediction by family composition in the segmentation system.....</i>	<i>63</i>

List of Equations

<i>Equation 1 - CIE ΔE_{2000}.</i>	21
<i>Equation 2 - Shannon entropy.</i>	22
<i>Equation 3 - Lower and Upper fences for mild outlier detection.</i>	34
<i>Equation 4 - Minimum sample size for SEM. (Christopher Westland, 2010)</i>	45
<i>Equation 5 - Cross-entropy loss function.</i>	54

List of abbreviations and definitions

Accumulation	Simulates larger batch size by accumulating gradients across individual batches.
AdamW	Optimiser algorithm with weight decay for regularisation.
ADE20K	Dataset with 150+ indoor/outdoor categories.
AI	Artificial Intelligence.
ASPP	Atrous Spatial Pyramid Pooling – allows multi-scale feature extraction.
Baidu	Chinese digital map and navigation service.
Batch	Group of images passed one at a time through a segmentation model.
BGR	OpenCV's default Blue-Green-Red channel order for images.
CamVid	Video-based street-scene dataset.
CE	Cross-entropy: Loss function used to train classification models.
CFA	Confirmatory Factor Analysis.
CFI	Comparative Fit Index.
CIELab*	Perceptually uniform colour space.
Cityscapes	Urban street-scene dataset.
CNN	Convolutional Neural Network.
COCO	Common Objects in Context (dataset).
CPU	Central Processing Unit
DCNN	Deep Convolutional Neural Network.
DeepLabV3	CNN architecture for segmentation, enhanced with atrous convolutions and ASPP.
EFA	Exploratory Factor Analysis.
Eigenvalues	Values used in factor analysis to determine factor importance.
Epoch	Complete pass through all training image batches.
FCN	Fully Convolutional Network.
GPU	Graphics Processing Unit
GSV	Google Street View
gtFine	Cityscapes ground-truth fine annotations.
GVI	Green View Index.
HSV	Hue-Saturation-Value colour space.
ImageNet	Organized global image database.
IoU	Intersection over Union – metric for evaluating segmentation accuracy.
IQR	Interquartile Range.
KMO	Kaiser-Meyer-Olkin measure of sampling adequacy.
Memory cache	Speeds up performance by storing frequently accessed data.
Mixed precision	Using 16-bit floats to save memory and speed up computation.
NIMBY	“Not in My Back Yard” effect.
Oblimin	Oblique rotation method in factor analysis.
OneCycleLR	Dynamic learning rate scheduler with single cycle of increase/decrease.
PAF	Principal Axis Factoring.

PCA	Principal Component Analysis.
Point-biserial	Special case of Pearson's correlation for binary x continuous variables.
Polychoric	Estimated correlation between ordinal variables.
POV	Point Of View
Pretrained weights	Model weights obtained from training on large datasets, reused for transfer learning.
PSPNet	Pyramid Scene Parsing Network.
r	Pearson's correlation coefficient.
RAM	Random-Access Memory
Residual-centring	Method to compute interaction terms while reducing multicollinearity.
RGB	Red-Green-Blue (colour space).
rho	Spearman's rank correlation coefficient.
RMSEA	Root Mean Square Error of Approximation.
Seed	Starting value for a pseudorandom number generator.
Segmentation	Computer vision technique - assigns a class label to every single pixel in an image.
SegNet	Convolutional network architecture for segmentation.
SEM	Structural Equation Modelling.
Shannon (H)	Measure of colour diversity in an image, normalised between 0 and 1.
spc_var	Per-class space variation between before and after images.
SRMR	Standardised Root Mean Square Residual.
Streetscore	A machine learning algorithm that predicts the perceived safety of a streetscape.
SVF	Sky View Factor: Metric to quantify openness of the sky above street.
t1	Baseline probability of preference estimate.
TLI	Tucker-Lewis index
U-Net	Convolutional network architecture for segmentation.
Varimax	Orthogonal rotation to maximize loading separation.
Weight decay	Regularisation technique that penalises large weights.
Weights	Represent the learned knowledge acquired from data during training
WLSMV	Weighted Least Squares Mean and Variance adjusted estimator.
ZenSVI	Platform combining street-level imagery with spatial indicators.
z-value	Indicates how many standard deviations a data point is from the mean

1 Introduction

1.1 Motivation

Climate concerns, social expectations and a desire for a healthier lifestyle are reshaping the way cities and streets are thought about. For a long time, urban design focused on vehicles. Roads were built wide, sidewalks were narrowed and green areas were often treated as afterthoughts. Today, that car-first model no longer fits in ([Alcoforado et al., 2009](#); [Ma et al., 2024](#)).

More than half of the world's population already live in cities ([González Mendoza et al., 2019](#)). Cities are also where most emissions, energy use and pollution come from, being an essential environment where strategies against climate change should be tackled ([Alcoforado et al., 2009](#)). Shifting street space away from cars and toward public transport and active modes is one of the most direct ways to act on these challenges. Streets that make the sustainable choices easier and more attractive avoid car-dominated layouts and tend to reinforce independence from private vehicles reducing pollution ([Gössling, 2020](#)). However, urban space distribution must go beyond technical plans and design guidelines since change only works if people support it.

When a city reallocates space for a new cycle lane or replaces parking spaces with trees and benches not only the performance is important but also whether people believe that the intervention has improved the built environment ([X. Li et al., 2015](#); [Loyola et al., 2023](#)). Many urban projects fail or face pushbacks due to a lack of evaluation on how people may perceive those changes before implementing them. For example, if a green intervention makes a street feel darker or more isolated, residents might oppose to it, even if it is environmentally friendly ([De Nadai et al., 2016](#); [Rashid et al., 2017](#)).

Studies show that perception is influenced by different visual and emotional attributes such as colours, trees, lighting and how busy a space feels ([Gorzaldini, 2016](#); [Y. Zhang et al., 2022](#); [Z. Zhang et al., 2021](#)). Some people feel more at ease in green streets, while others worry about losing parking or driving space. Gender, age and personal habits also play a role. Women, for example, often report feeling less safe in the same public spaces as men ([Cui et al., 2023](#); [Rashid et al., 2017](#)). Also, younger generations are starting to shift away from car culture, preferring to walk, cycle or take public transport ([Merlo et al., 2025](#); [Park & Garcia, 2019](#)).

Despite this, most planning tools still do not account for how people feel about street changes. Evaluation tends to rely on technical drawings, traffic counts or before and after statistics. However, emotional responses, perception of comfort and visual appeal are rarely measured or estimated in a quantitative manner ([Larkin et al., 2021](#); [Moreno-Vera et al., 2021](#)). Thus, understanding how people experience using streets is essential to plan effective sustainable interventions that people use and accept.

Emerging tools such as Street View imagery, deep learning models and perception surveys are recently being used to study how people perceive and experience urban environments at a much bigger scale ([Dubey et al., 2016](#); [Ito et al., 2025](#); [Yao et al., 2019](#)). Instead of only asking people what they think, it is possible to collect thousands of reactions and correlate them with image-based changes. Recent literature, still extremely limited, has focused on how visual differences (e.g., adding vegetation or removing cars) affect perceptions of safety, liveliness or beauty ([C. Chen et al., 2022](#); [Wu et al., 2024](#); [W. Zhao](#)

et al., 2024)

In this context, this thesis is motivated by the potential to help fill the gap between technical design and public perception. The work is based on evaluating perceptions of before and after images of street interventions through an online survey, classifying their visual changes through image recognition and linking those changes with how people feel about the interventions. This work contributes to estimating the kind of visual transformations that are most likely to be welcomed or rejected by the community (T. Li et al., 2020; Sádaba et al., 2024; Verma et al., 2020).

In short, this work aims to make sustainable urban planning not just smarter, but more human-centric. Streets are not just infrastructure; they are public spaces where daily life happens (Gehl, 1936). To fulfil the need for public preference in street interventions, urban planning needs to add, to the engineer and policymaker perspective, the emotions of the citizens who cycle, walk, rest and live there.

1.2 Problem definition and research questions

Public preference often falls behind the city's growth and evolution. Many street projects designed to improve safety, sustainability or accessibility end up contested or underused. While planners focus on technical solutions, the way people feel about those changes is still poorly understood (Alcoforado et al., 2009; González Mendoza et al., 2019; Webster, 1998). Streets are as much infrastructures as they are social spaces. Therefore, evaluation methods should not overlook perception entirely.

Current tools rarely capture how residents interpret visual changes like added greenery, altered road width or new infrastructure. Global platforms such as Place Pulse, a tool for collecting and analysing public perceptions of urban environments by asking users to compare two randomly selected street images based on safety, beauty and liveliness, offer some insight into perception, but these models tend to generalize across vastly different cultures, geographies and demographic groups (Dubey et al., 2016; Muller et al., 2022; Yao et al., 2019). Without the ability to adapt perception models to local realities, design decisions risk being misaligned with public sentiment.

The visual human perspective plays a central role in how people evaluate their surroundings. Changes in the colour, structure and greenery of a street can influence emotions, perceived safety and even behavioural choices like walking or cycling (Cui et al., 2023; De Nadai et al., 2016; Larkin et al., 2021; Z. Zhang et al., 2021; W. Zhao et al., 2024). Yet few urban studies have explored this link in a way that is both scalable and sensitive to context. The result is a gap between the physical transformation of a space and the psychological experience it creates.

Efforts to study perception are also held back by limitations in methodology. Traditional surveys and audits provide fragmented snapshots, often lacking the spatial or visual richness needed to capture the full street experience (N. He & Li, 2021; Verma et al., 2020). Newer approaches using deep learning and semantic segmentation are promising, but their outputs are not always interpretable in terms of emotional or perceptual outcomes (Hao et al., 2024; T. Li et al., 2020). Many models identify visual features but cannot explain how those features translate into public support or rejection of an intervention. Other studies discuss a complete replacement of street design by generative AI models such as UrbanistAI, emphasizing the *“(...) need for further studies to confirm hypotheses, understand psychological responses and impacts on critical thinking and develop strategies for ethically and effectively*

incorporating these tools into public participation events (...)" (Valença et al., 2025).

Another challenge lies in how perception varies across groups. Users do not all equally experience safety, beauty and comfort. Age, gender and familiarity with a place can shift how people read a space visually (Atiqul Haq et al., 2021; Cui et al., 2023; De Nadai et al., 2016; Merlo et al., 2025). A single change like planting tall vegetation might make some feel safer while making others more anxious (Rashid et al., 2017; Y. Zhang et al., 2022). Understanding these nuances is essential to ensure inclusive urban design.

The main objective of this work was to explore how visible changes, such as more greenery, less asphalt or brighter colours, affect the way people feel about different street interventions (e.g., reallocating a traffic lane to a bus lane). To achieve this, the project analyses before and after images of urban interventions, detecting what changed and linking those differences to how people perceive the importance of comfort, safety or liveability. Colour analysis and semantic segmentation help identify both structural and detail shifts, while survey responses and demographic data add public perceptions and context to those reactions. In terms of policy, this work offers a scalable and adaptable image-based framework to measure and interpret perception in urban environments, especially in contexts where traditional participatory tools are hard to deploy at a large scale. To support this aim, the following research questions are proposed:

Question 1: How can emotional responses to street environments be quantified and integrated into evidence-based approaches to urban design?

Question 2: Can exploratory predictive tools anticipate public preference to street redesigns before implementation?

By addressing these questions, this dissertation sought to connect physical design and human emotional responses, by creating a tool to identify the attributes that contribute to public preference towards distinct kinds of urban street distribution. Rather than offering a one-size-fits-all answer, the goal was to create an adaptable methodology that could predict which visual transformations were more likely to be preferred depending on the population being evaluated. This dissertation combined survey data, corrected semantic segmentation, perceptual colour metrics, univariate/bivariate analyses, polychoric approaches and SEM-based modelling, operationalising a predictive tool capable of simulating public reactions to urban interventions. The framework was also applied to a real-world case study, demonstrating its practical implementation and the ability to anticipate public preference before project execution. By using the same method, urban designers could make more informed decisions and build spaces that are welcomed by the communities they serve.

1.3 Dissertation structure

After the introduction to this dissertation's theme, research questions and surrounding motivation, this document is structured to provide a logical progression from the existing literature to the methodological design choices, followed by the results and conclusions and ends with the discussion on contributions and limitations of the framework defined and suggestions for future work.

This progression starts with chapter 2 State of the art, which covers previous research on public perceptions of street design, image-based urban sensing methods, colour metrics, computer vision approaches, methodological considerations for temporal image comparisons and relevant

psychometrics and statistical modelling techniques. This chapter concludes with an identification of research gaps that motivate the subsequent methodology.

This methodology is then described in chapter 3 Methodology, which begins with descriptions of survey data and image collection, preprocessing and colour analysis. Afterwards, the development and inference of the segmentation model are detailed, along with the measurement of variation between images. This chapter also explains the statistical analyses conducted, including univariate and bivariate analyses, polychoric exploratory and confirmatory factor analysis on perception variables and structural equation modelling (base and moderation models). Finally, the predictive tool developed to estimate public responses based on image analyses is introduced and its application to a real-world case study is presented to demonstrate a practical implementation of the anticipation of public preference.

Then, chapter 4 Results and discussion discusses the outcomes of image-derived colour and segmentation analysis, followed by survey response patterns and ends with the application of the predictive tool in practical scenarios. This chapter also addresses the implications of the findings and the limitations of the methodology described.

Lastly, chapter 5 Limitations and future work, summarises the main insights, contributions and directions for future research, concluding with final reflections on the dissertation.

2 State of the art

The state of the art situates this dissertation within a multidisciplinary field that spans urban design, environmental and perception-based psychology, computer vision and colour metrics and statistical modelling. It begins by reviewing how perceptions and preferences shape the success of street interventions then turns to the methodological developments that have enabled researchers to measure these perceptions more systematically, from image-based sensing to psychometric modelling.

At the same time, this chapter highlights the limitations of existing approaches, including dataset biases, issues of cultural generalisability and gaps in connecting perception research with planning practice. By synthesising these contributions and challenges, the State of the Art both defines the intellectual foundations of this work and clarifies the research gaps that the thesis addresses through its focus on before and after workflow, perceptual colour metrics and decision-support modelling.

2.1 Perceptions and public preference on street design

People's subjective reactions to street environments, such as how safe, comfortable or attractive they feel, are not marginal considerations but essential determinants of how those environments are used and preferred. A street that is technically safe on paper may still be avoided if it feels unsafe, dull or uninviting. Conversely, interventions such as pedestrianisation or greening depend on public preference to succeed ([Muller et al., 2022](#); [Yao et al., 2019](#)).

This thesis evolves around this gap between functional performance metrics and human preference, arguing that perceptions form a critical layer in understanding and predicting urban behaviour. This explains why the methodological design of this research focuses on the importance people assign to perceptual dimensions rather than on direct reported feelings. By doing so, the analysis could capture what citizens care about most in evaluating interventions, which is a more policy-relevant perspective.

These perceptions are not uniform. They are multidimensional, span from safety, comfort, usefulness order, security, stress and liveliness and they are socially distributed across gender, age and personal experience ([Cui et al., 2023](#); [De Nadai et al., 2016](#); [Merlo et al., 2025](#)). Recognising these complexities provides the necessary foundation for linking perceptual theory to the analytical framework of this thesis.

The literature has consistently shown that perceptual reactions to street design are influenced by both physical attributes and socio-demographic characteristics. Several studies demonstrate that greenery plays a central role, often increasing perceived safety and comfort, though the type and placement of vegetation matter. Research using crowdsourced image ratings found that tall vegetation above the horizon line tends to enhance perceived safety, while vegetation below the horizon can have more ambiguous effects ([Dubey et al., 2016](#)).

Similarly, studies in Chinese cities have shown that higher levels of visible greenery reduce fear of crime, though this relationship is mediated by the presence of incivilities such as graffiti or signs of neglect ([Hao et al., 2024](#); [X. Li et al., 2015](#)). Other work conducted in controlled virtual environments highlights the impact of traffic lanes, vegetation and lighting conditions on perceived safety at crossings, showing that fewer lanes, greener edges and warmer lighting make pedestrians feel more secure ([Duncan et al., 2016](#); [W. Zhao et al., 2024](#)). These findings collectively justify the methodological decision to quantify

greenery and visual elements in the image analysis phase of this research and to examine their relation to perceptions such as safety and comfort.

Beyond greenery, other visual and physical factors exert significant influence. Lighting conditions and colour tones contribute to whether a space feels welcoming or threatening (Dabbagh, 2019; N. He & Li, 2021), while open skies and wider sidewalks have been associated with higher comfort and liveliness (Y. Zhang et al., 2022).

Demographic factors also play an important role. Women have been found to be more sensitive to lighting quality and the presence of people, while older adults are more attentive to sidewalk conditions and maintenance (Park & Garcia, 2019; Rashid et al., 2017).

A recent global survey of urban visual perception demonstrated that demographic variation extends beyond simple categories, as personality traits and cultural background can significantly alter how a space is judged (Quintana et al., 2025). These results support the methodological choice in this thesis to include demographic moderators in the analysis and to avoid treating perceptions as universal truths detached from the identities of respondents.

2.2 Image-based urban sensing and participatory evaluation methods

The incorporation of image-based sensing and digital participation platforms has profoundly reshaped how urban perception can be studied. Traditionally, assessments of the built environment relied on field surveys, interviews, focus groups or systematic audits conducted by trained observers. These methods were rich in qualitative insight but limited in scale and were often resource intensive (Cavaco et al., 2015; Crouse, 2004; Webster, 1998). These approaches provided valuable context but were constrained by the number of sites that could be covered and the demographic biases of participants willing to engage in time consuming interviews.

The research of large-scale visual datasets, particularly those derived from street-level imagery, has provided a way to scale perception studies while maintaining a realistic representation of the environments being studied (N. He & Li, 2021; Muller et al., 2022). This dissertation situates itself within this transition by combining before and after images of interventions with survey-based evaluations, leveraging the scalability and visual realism of newly achievable methods while retaining the contextual sensitivity of traditional participation.

One of the key innovations has been the use of crowdsourced image rating platforms such as Place Pulse, which collected millions of pairwise comparisons of urban images on dimensions like safety, wealth and liveliness (Dubey et al., 2016). By mobilising large numbers of online participants, Place Pulse demonstrated the feasibility of mapping subjective perception across cities and even countries, for comparative research.

However, subsequent analyses showed that such platforms can oversimplify urban complexity and suffer from respondent bias, as online participants often skew toward younger and more digitally literate populations (Larkin et al., 2021; Moreno-Vera et al., 2021). Nonetheless, Place Pulse and similar platforms enable the research around a public participation methodological shift based on photographic stimuli. By using image pair surveys, this thesis draws directly from these precedents but adapts them to a before and after workflow that permits the evaluation of concrete street interventions.

Street-level imagery from platforms such as Google Street View (GSV), Baidu and Tencent has become a standard data source for urban analysis, offering consistent perspectives at pedestrian eye level (N. He & Li, 2021; Ito et al., 2025). Researchers have used this data to quantify greenery via the Green View Index (GVI), measure openness through the Sky View Factor (SVF) and assess walkability and aesthetic qualities without needing extensive fieldwork (Hao et al., 2024; X. Li et al., 2015).

However, the use of GSV comes with strict copyright restrictions, making large-scale automated analysis problematic (Dubey et al., 2016). For this reason, the present dissertation uses open-access alternatives such as Mapillary and KartaView, complemented by project-sourced images and personal photographs. This choice allows for a reproducible and ethically sound methodology that avoids legal complications while ensuring that image data remain tied to the specific interventions under study.

Recent advances in computer vision have also further expanded the potential of image-based urban sensing. Automated semantic segmentation enables the extraction of features such as greenery, built structures or sky openness, which can then be linked to perceptual outcomes measured via surveys (Dong et al., 2020; T. Li et al., 2020; W. Zhao et al., 2024). Platforms such as ZenSVI combine street-level imagery with spatial indicators to produce objective estimates of street vitality based on elements like pedestrian presence or vegetation (Ito et al., 2025). These approaches represent a shift from purely subjective and quantitative assessments towards models that combine human responses with machine-derived predictors. The methodology of this dissertation integrates segmentation results with survey-based importance ratings, positioning it at the intersection of subjective perception research and automated scene analysis.

The framing and resolution of imagery strongly influence the validity of perception studies. Experimental work shows that differences in angle, distance or lighting can alter participants' responses, even when the underlying environment is the same (Dabbagh, 2019; Y. Zhang et al., 2022). Seasonal variations also significantly affect vegetation and lighting, thereby influencing perception scores in ways that are not always linked to the intervention itself (Hao et al., 2024). Renderings and photo simulations allow researchers to control such variables and to isolate specific features like greenery or building design. Global datasets such as Place Pulse also allow researchers to make broad generalisations, but they often obscure local specificities and cultural differences that are crucial for planning practice (Muller et al., 2022).

This dissertation evaluates real photographs before and after specific interventions to allow the use of the survey data collected and described by (Lopes, 2025). At the same time, it acknowledges the need for systematic consistency in framing and viewpoint to ensure comparability, reflecting best practices outlined in literature (Hao et al., 2024; Tang & Long, 2019).

2.3 Colour and visual metrics used in perception research

Colour plays a central role in how people perceive and evaluate urban environments and feelings of safety, comfort, beauty and liveliness. Unlike physical features such as road width or vegetation, colour operates at a different level of emotional and cognitive response that individuals have when interacting with streetscapes (Dabbagh, 2019; Gorzaldini, 2016).

For this reason, colour analysis deserves a dedicated review within the state of the art, particularly

because the methodology of this thesis employs perceptually grounded metrics such as ΔE_{2000} , hue-saturation-value (HSV) binning and Shannon entropy to quantify visual variation. These methods allow for a complete analysis of how urban interventions alter the perceptual qualities of a space. The decision to use perceptual colour models rather than raw red-green-blue (RGB) values reflects both evidence from environmental psychology and findings from colour science, which demonstrate that RGB is poorly aligned with human perception, whereas colour spaces such as CIE Lab and its derivative ΔE_{2000} were explicitly designed to approximate how differences are experienced by observers (Luo et al., 2001). By anchoring analysis in perceptual metrics, this thesis ensures that visual changes are quantified in ways that reflect actual human sensitivity to colour differences.

The literature shows that colour features influence urban perception in multiple ways. Studies have found that brightness and warm-toned lighting increase feelings of safety and attractiveness, whereas low-light or harsh lighting tends to decrease comfort (N. He & Li, 2021; Park & Garcia, 2019). Similarly, vegetation greenness, often measured via indices such as the Green View Index (GVI), has been repeatedly associated with restorative experiences, perceived safety and higher liveability scores (Hao et al., 2024; X. Li et al., 2015).

Colour composition also matters, since harmonious or diverse palettes influence perceived aesthetics, with entropy measures of palette complexity being linked to perceptions of vibrancy or disorder (Yu et al., 2024). These findings base the methodological choice in this dissertation to compute Shannon entropy as a variable for colour diversity, thereby allowing the measurement of whether interventions create streetscapes that feel visually engaging or, alternatively, cluttered.

Several controlled experiments demonstrate how changes in colour and brightness alter perceptual judgments. For example, photo-simulation studies manipulating vegetation or colour tones found that greener or brighter versions of the same street were consistently rated as safer and more inviting (Y. Zhang et al., 2022; Z. Zhang et al., 2021).

Laboratory research in colour science also demonstrates that small ΔE_{2000} differences, often as low as 1 or 2 units, are perceptible under controlled conditions (Luo et al., 2001). However, when applied to uncontrolled urban images, noise from shadows, weather and seasonal variation can obscure or exaggerate differences. This justifies the empirical ΔE_{2000} threshold adopted in this thesis, which balances the theoretical perceptibility of small colour shifts with the variability of real-world imagery. In other words, while colour science sets the lower bounds of human sensitivity, empirical testing is needed to calibrate thresholds for street-level analysis, ensuring that metrics capture meaningful perceptual differences rather than trivial pixel-level fluctuations and noise.

The use of HSV bins has also gained traction because the separation of hue, saturation and brightness reflects distinct perceptual dimensions. Hue captures the qualitative aspect of colour, saturation relates to intensity and brightness corresponds to luminance, each of which has been shown to shape psychological responses differently (Dabbagh, 2019; Gorzadini, 2016). For instance, high saturation has been associated with vibrancy and attractiveness, while high brightness correlates with perceptions of safety and openness (Loyola et al., 2023; Verma et al., 2020). By applying HSV analyses to street imagery, this thesis aligns its approach with empirical evidence that these dimensions are not interchangeable and contribute uniquely to perceptual judgments.

2.4 Computer vision for street-scene analysis

Semantic segmentation has become one of the most powerful tools for analysing urban imagery, enabling every pixel of a street scene to be assigned to a semantic class such as road, building, vegetation or pedestrian (T. Li et al., 2020). For perception studies, this pixel-wise detail is crucial because it makes it possible to quantify the relative share of elements like greenery, obstacles or road lanes. This development made the present methodology possible, where the selection of DeepLabV3 with a ResNet-101 backbone and fine-tuning on the Cityscapes dataset provided a balance between segmentation accuracy and computational efficiency. The trade-offs of this choice are central to the overall workflow: Cityscapes offers high-quality labels for central European street scenes, which align with the studied contexts, but its label scheme is incomplete (no bicycle or bus lanes). Recognising these limitations motivated the incorporation of a manual correction phase in the methodology, ensuring that misclassified or missing objects could be reconciled prior to the before and after comparison analysis.

The development of deep learning architectures marked a major turning point in segmentation research. Fully Convolutional Networks (FCNs) first allowed pixel-level predictions across entire images, replacing patch classification and reducing the need for handcrafted features (Long et al., 2015). Subsequent models introduced architectural innovations such as U-Net improving segmentation in small detail datasets (Ronneberger et al., 2015), SegNet optimising memory use (Badrinarayanan et al., 2015), PSPNet advancing in multiscale context (H. Zhao et al., 2017) and DeepLab series introducing atrous convolutions for flexible receptive fields (L.-C. Chen, Papandreou, Kokkinos, et al., 2017). Lastly, DeepLabV3+ combined these advances with an encoder–decoder structure, improving delineation of object boundaries and thin structures, such as street poles or sidewalk edges (L.-C. Chen et al., 2018).

Among these options, DeepLabV3 with ResNet backbones has consistently achieved high mean Intersection over Union (IoU) scores on benchmarks like Cityscapes and COCO, making it a sensible choice for this dissertation where both fine-grained accuracy and computational feasibility in regular workstations was necessary (Cordts et al., 2016; Lin et al., 2015).

Datasets are also decisive in segmentation quality. The Cityscapes gtFine version of the dataset remains a standard for urban environments, providing pixel-level annotations for thirty-four classes across thousands of finely labelled central European street scenes (Cordts et al., 2016). However, it is geographically and culturally specific, which can limit applicability to other regions. ADE20K, with more than 150 categories across indoor and outdoor contexts, offers broader coverage but less depth in street specific classes (Zhou et al., 2018). COCO adds object diversity but emphasises instance segmentation rather than dense street layouts (Lin et al., 2015). CamVid supplies video-based annotations useful for temporal continuity but at lower resolution (Brostow et al., 2009).

Each dataset has strengths and weaknesses and transfer learning strategies are often employed to adapt pretrained networks to the new domains (Yosinski et al., 2014). In this thesis, Cityscapes was chosen to tune a pretrained model (from COCO) for its direct relevance to street-level analysis, while augmentation strategies such as random crops, flips and brightness adjustments were introduced to mitigate overfitting and improve robustness across different lighting, weather and perspectives.

Evaluation metrics are not only used to interpret model performance, but also to teach it. These depend on the study's objectives and enable better model learning. Mean IoU, for example, measures the overlap between predicted and ground-truth pixels across classes. However, not all classes are equally relevant to perception studies. For this reason, mean relevant IoU, which can focus on road, sidewalk, vegetation, sky and building classes, was chosen as a more informative measure for the objectives defined in chapter 1.2 Problem definition and research questions.

2.5 Methodological concerns for before and after image comparison

Comparing before and after images of the same location is one of the most direct ways to evaluate the impact of urban interventions, yet it is also one of the most technically and methodologically demanding, which might justify its scarcity in literature. Minor changes in camera viewpoint, lighting conditions, seasonal variations or image resolution can introduce artefacts that risk being misinterpreted as intervention effects rather than environmental noise (Dubey et al., 2016; Hao et al., 2024).

These concerns were recognized in this thesis, since the analytic framework depends on quantifying changes in pixel allocations and colour distributions between paired images. To ensure consistency, all images were resized to a standard dimension, aligned as closely as possible through cropping and converted into perceptually meaningful colour spaces prior to analysis. The decision to use a single before and after pair per intervention avoids double counting but raises the risk of representativeness.

The literature provides several best practices for handling before and after image comparisons. Alignment is essential, as demonstrated by studies using time-series street view imagery to evaluate façade evolution, where shifts in camera perspective can otherwise obscure or exaggerate changes (Dubey et al., 2016). Researchers recommend consistent framing and scale adjustments to minimise these discrepancies (Tang & Long, 2019).

Seasonality is another recurrent issue, because interventions captured in full-leaf summer conditions often appear greener and more inviting than winter counterparts, regardless of the physical changes introduced (Hao et al., 2024). This has led to research around either seasonal controls or statistical corrections in image datasets.

Illumination also matters, as shadows and glare alter perceived safety and cleanliness, with some studies suggesting the inclusion of lighting metadata or brightness normalisation in preprocessing workflows (X. Li et al., 2015; W. Zhao et al., 2024). The present thesis mitigated these risks by adopting perceptual colour metrics (ΔE_{2000} , HSV distributions), less sensitive to absolute illumination than raw RGB colour space.

Preprocessing choices also carry implications for measurement error. For example, choosing a single before and after pair reduces the noise from multiple non-identical captures but risks bias if the chosen pair is not representative of typical conditions. Other studies have sought to average across multiple captures or to simulate after-images through photo editing or rendering, but these approaches bring their own challenges of realism and respondent trust (Y. Zhang et al., 2022; Z. Zhang et al., 2021).

This dissertation took a pragmatic route, privileging real-world pairs while acknowledging that some residual noise remains. Likewise, the threshold for ΔE_{2000} was derived from empirical tests rather than laboratory psychophysics, recognising that uncontrolled street imagery introduces variability beyond

what small “*just noticeable difference*” thresholds imply in lab conditions ([Sharma et al., 2005](#)).

2.6 Psychometrics and statistical modelling

Bridging the analysis of visual street features with the measurement of subjective perceptions requires psychometric and statistical frameworks. Unlike computer vision tasks, which directly output pixel classes or colour values, perception research deals with latent constructs. These are abstract dimensions which cannot be measured directly.

In this dissertation, latent factors were constructed through survey items, which were treated as ordinal Likert-scale indicators. Factor analysis methods were used to uncover the underlying dimensions of these survey declared emotions, leading to a three-factor solution interpreted as the perceived importance of functionality, vulnerability and habitability. This measurement model was then integrated into a structural equation modelling (SEM) framework, which enabled the testing of mediated and moderated pathways from visual street features to public preference.

The use of psychometrics in urban perception research builds on a long tradition in social sciences. Exploratory factor analysis (EFA) has been widely used to identify latent structures in environmental attitudes, neighbourhood satisfaction and walkability perceptions ([Brown, 2015](#)). Confirmatory factor analysis (CFA) allows researchers to validate these structures against theory and data, testing model fit through indices such as RMSEA, CFI and TLI ([Hu & Bentler, 1999](#); [Kline, 2016](#)).

In studies of urban environments, these approaches have revealed that perceptions are multi-dimensional, often loading on separate but correlated factors rather than collapsing into a single latent construct ([Arellana et al., 2019](#); [Rajapaksa et al., 2018](#)). Such evidence justifies the decision to distinguish the importance of functionality, vulnerability and habitability as separate dimensions, instead of grouping all indicators into one latent construct, reflecting both theoretical reasoning and empirical data.

Structural equation modelling has also become increasingly relevant in urban studies because it allows the integration of measurement models with causal hypotheses. For instance, SEM has been applied to assess how environmental perceptions interact with socio-demographic factors, showing that gender, age and cultural background can moderate pathways from objective features to subjective evaluations ([De Nadai et al., 2016](#); [González Mendoza et al., 2019](#)).

These applications highlight some of the advantages of SEM over simple regression or machine learning approaches such as explicitly accounting for measurement error and allowing direct, indirect and moderated effects to be estimated. In this thesis, SEM was used in precisely this way, linking image-derived measures such as vegetation pixel-share and brightness metrics to latent perception factors and then to stated preferences for each intervention.

The statistical approaches chosen also reflect ongoing debates in the field. Treating Likert items as ordinal rather than continuous has been recommended for improving accuracy in small to medium samples, particularly when response distributions are skewed ([Remtulla et al., 2012](#)). Polychoric correlations better capture the association structure between ordinal items than Pearson correlations, which can underestimate loadings. Similarly, WLSMV estimators have been shown to produce more reliable standard errors and fit indices in ordinal CFA and SEM than maximum likelihood under conditions of categorical data ([C. H. Li, 2016](#)).

Additionally, principal component analysis (PCA) was used to reduce correlated colour metrics into interpretable dimensions, following approaches in environmental psychology where perceptual complexity must be captured in a parsimonious manner (Dabbagh, 2019; Gorzaldini, 2016). Residual-centred interactions were chosen for moderation tests to avoid multicollinearity and improve interpretability, a strategy supported by methodological literature as well (Geldhof et al., 2013).

The psychometric and statistical modelling framework adopted in this thesis reflects a trade-off between rigour, interpretability and feasibility. By anchoring its latent constructs in both theory and empirical data, employing robust estimators for ordinal indicators and carefully managing moderation analysis, the methodology positions itself as a step toward more transparent and robust models of how urban visual features shape public perception and preference.

2.7 Summary and Research Gaps

Urban perception research has grown over the past two decades, demonstrating that how streets are experienced plays a decisive role in shaping public behaviour, attitudes and support for interventions. Studies consistently show that safety, comfort, aesthetics and liveliness are core dimensions of perceived street quality and that these perceptions mediate the effectiveness of policies aimed at shifting mobility patterns (Duncan et al., 2016; Muller et al., 2022; Rajapaksa et al., 2018).

This recognition has spurred the development of new methodologies combining computer vision, psychometrics and participatory approaches. Nevertheless, the reviewed literature highlights persistent gaps in generalisability, methodological transparency and integration with planning practice, which this dissertation addresses through a novel methodology that combines before and after imagery, corrected semantic segmentation, perceptual colour metrics, survey-derived demographic and habit descriptors and a more interpretable approach on structural equation modelling.

Existing studies often focus on single cities or specific urban morphologies, limiting external applicability. For example, widely used datasets such as Cityscapes and COCO are heavily Eurocentric and traffic-oriented, which constrains the diversity of represented built forms and vegetation types (Cordts et al., 2016; Neuhold et al., n.d.). Likewise, global perception platforms such as Place Pulse achieved scalability but fell short in cultural transferability, as perceptions of safety or attractiveness collected in Boston or Shanghai do not always map onto other regions (Dubey et al., 2016; X. Li et al., 2015). This reliance on narrow training and survey data risks encoding structural bias into models, reducing their inclusivity when applied elsewhere. By contrast, the present thesis builds its analysis on real before and after interventions and integrates demographic moderation in an automated statistical way, thereby grounding findings in relevant contexts while providing a transferable workflow that can be adapted to other cities by rerunning surveys and image analyses.

Another research gap concerns the lack of longitudinal and intervention-specific studies. Most perception research compares static conditions or correlates single images with subjective ratings rather than directly evaluating design changes. Only a handful of projects have attempted before and after comparisons (Lopes, 2025) and even fewer have done so with controlled image alignment (Wu et al., 2024). This dissertation responds to that need by explicitly adopting the before and after survey designed by (Lopes, 2025) anchored in photographic records of street transformations, which makes it possible to

quantify perceptual changes attributable to transformations rather than to seasonal variation or random image selection. To further address residual risks of lighting and seasonal bias, colour analyses were standardised using perceptual distance measures such as ΔE_{2000} and vegetation results were cautiously interpreted with seasonal information (Hao et al., 2024; Tang & Long, 2019).

Colour is another underdeveloped dimension in literature. While greenery has been widely studied as a descriptor for positive affect (X. Li et al., 2015; Y. Zhang et al., 2022), more nuanced metrics of brightness, saturation, hue presence and entropy have rarely been used in perception research (C. Chen et al., 2022). Yet, controlled psychology experiments show that such variables significantly affect mood, attention and aesthetic judgement (Yu et al., 2024).

By operationalising ΔE_{2000} perceptual distances alongside HSV brightness and saturation measures, this thesis introduces colour into urban perception modelling in a comprehensive approach that is reproducible, scalable and interpretable. This methodological innovation makes it possible to capture subtler aesthetic changes than the binary green versus non-green distinction used in earlier work.

The question of representativeness also remains a major limitation. Crowdsourced perception surveys tend to be skewed toward younger and technologically literate populations (Ito et al., 2025), while traditional household surveys capture more diversity but are resource-intensive and not scalable. Moreover, demographic moderators are often neglected, even though research demonstrates that gender, age and income systematically influence perceptions of safety, vulnerability and comfort (Cui et al., 2023; Rajapaksa et al., 2018). This dissertation addresses this gap by explicitly modelling demographic effects within its SEM framework in an automated form, thereby identifying not only average tendencies but also group-specific differences in how interventions are perceived.

Bridging perception research with planning practice is another area where progress has been scarce. Tools such as Streetscore provide relative rankings of urban safety but lack interpretability and produce little actionable guidance for municipal planners (Dubey et al., 2016). Recent calls emphasise the need for decision-support tools that connect predictive models with planning interfaces and communicate uncertainty transparently (Pereira Martins, 2021; Sádaba et al., 2024). The predictive prototype developed in this thesis represents a response to these calls, by structuring inputs and outputs in a modular way that can be integrated into municipal platforms like Lisbon's NaMinhaRuaLX and further inform policy and decision making. Unlike forecasting engines that claim precision, this tool is explicitly positioned as a prioritisation instrument, highlighting which interventions are more likely to be preferred without overstating certainty.

Image derived descriptors and statistical methods need further research as well. Semantic segmentation models, even when corrected, represent another branch of the limitations found in literature, impeding the use of new technologies in urban planning's decision processes. They inherit biases from their training datasets and may systematically misclassify features that matter for perception, such as small-scale greenery or signage (Dong et al., 2020; Yu et al., 2024). Colour analyses are vulnerable to lighting variability, which complicates comparisons across different times of day or year. Lastly, SEM approaches require large samples for stability, yet urban perception surveys often operate with limited recruitment, leading to a trade-off between model complexity and reliability (Kline, 2016).

Furthermore, the widespread reliance on Google Street View (GSV) has raised copyright and ethical concerns while researching for this dissertation, since Google's strict rules protect their editing capabilities by prohibiting several use cases of their street perspective images, especially when workflows require screenshots or offline versions of the data. By deliberately avoiding copyrighted imagery, employing intervention photographs found through multiple public sources and designing transparent correction and validation steps, this thesis situates itself within an ethical framework while acknowledging the inherent limits of such safeguards.

The state-of-the-art reveals that urban perception research has achieved important advances in linking visual attributes to subjective experience, but it still struggles with generalisability, representativeness and integration into planning practice. This dissertation contributes with four innovations: a before and after image workflow that directly quantifies the transformations due to specific interventions; the incorporation of perceptual colour metrics into preference modelling; the automation of SEM's interaction structure based on statistical results for moderation relevance, which allows for wider reproducibility across different communities and the design of a predictive decision support prototype that balances methodological rigour with usability and interpretability for stakeholders.

By addressing these gaps, the research establishes both methodological contributions and practical tools that can help urban planners make more inclusive, evidence-based and publicly preferred design choices.

3 Methodology

3.1 Methodological framework

The proposed methodology integrated computer vision, colour perception metrics and statistical modelling to evaluate how street redesigns influence public perception. The workflow, illustrated in Figure 1, progressed from data collection to predictive modelling and validation.

It began with image and survey data gathering as pairs of before and after images of real street interventions were collected to visually document changes in the built environment. Interactive street view versions of these were presented to survey participants, who indicated which scenario they preferred and assessed the extent to which specific emotions such as comfort, stress, security, safety order, usefulness and liveability influenced their choice. This procedure linked perceptual reactions and concrete visual transformations, allowing emotional responses to later be quantitatively related to measurable design changes. In addition, respondents provided demographic information (e.g., age, gender, residence) and commuting habits (e.g., main commute mode, car ownership, access to parking). These contextual variables were collected to account for the differences in perceptions derived from users' backgrounds and mobility behaviours, ensuring that the methodology reflected the diverse perspectives among the respondents.

Using the gathered image pairs, two forms of visual analysis were performed. Colour metrics quantified global visual properties such as brightness, saturation and hue distribution while semantic segmentation isolated specific street components including road space, vegetation, sidewalks and cycling infrastructure. Since convolutional neural networks are prone to misclassify elements when trained on limited data ([Shorten & Khoshgoftaar, 2019](#)), a manual correction step was included to efficiently refine mislabelled regions, as suggested by [Reza \(n.d.\)](#). The resulting colour values and corrected segmentations enabled the calculation of proportional variations between before and after images, providing a structured measure of physical change for each pair.

All visual and perceptual indicators were then explored through a preliminary statistical analysis including univariate and bivariate approaches, followed by exploratory and confirmatory polychoric factor analyses. These steps identified significant predictors, minimised redundancy and derived latent constructs that better represented the underlying perceptual dimensions assessed in the modelling and predictive phases ([Hair JR. et al., n.d.](#)).

Subsequently, the integrated dataset was passed through structural equation modelling, allowing the investigation of direct, mediated and moderated effects between quantified design features and reported perceptions. From these results, a predictive tool was developed by translating the coefficients into linear predictors, with a shrink calibration procedure applied to maintain adaptability when estimating expected public reactions with new interventions or respondents.

Finally, a validation case study applied the predictive system to generated hypothetical scenarios, simulating alternative street configurations and evaluating their projected public preference using the same methodological pipeline. This stage demonstrated the framework's capacity to translate theoretical relationships into a practical decision-support tool for evidence-based urban design.

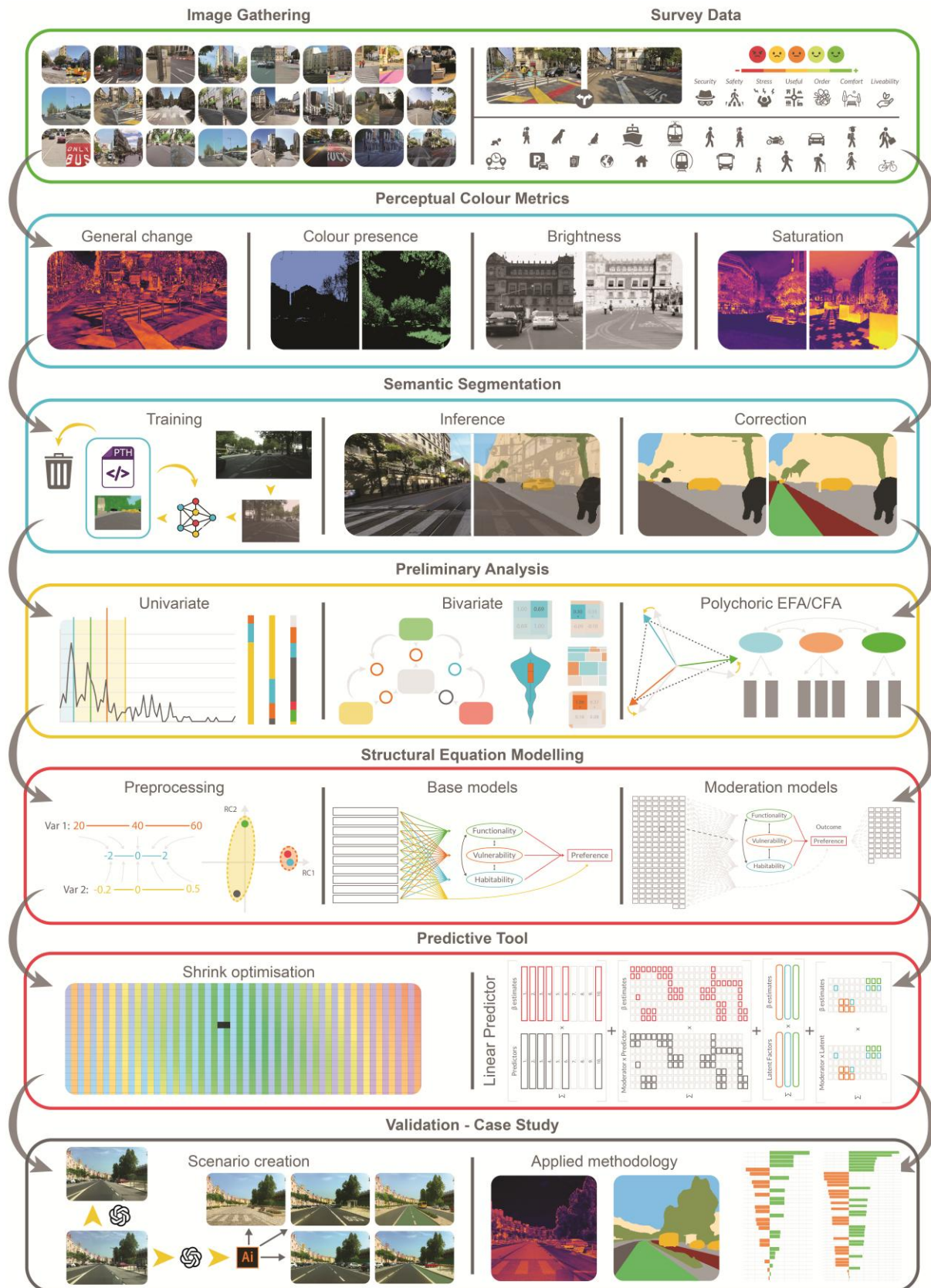


Figure 1 - Illustrated summary of the methodology.

3.2 Survey data

The survey data used in this dissertation was originally collected by [Lopes \(2025\)](#) as part of a masters degree dissertation. An online questionnaire was distributed to a general audience, with the objective of assessing how respondents perceive and evaluate urban street transformations when presented with before and after images of real interventions. The survey followed a structured sequence that combined image-based evaluations with a familiarity question, demographic and commuting information.

Each respondent was shown eight different interventions in the form of before and after image pairs, as exemplified in Figure 2.

*Please select the urban environment you **prefer**.

Choose one of the following answers

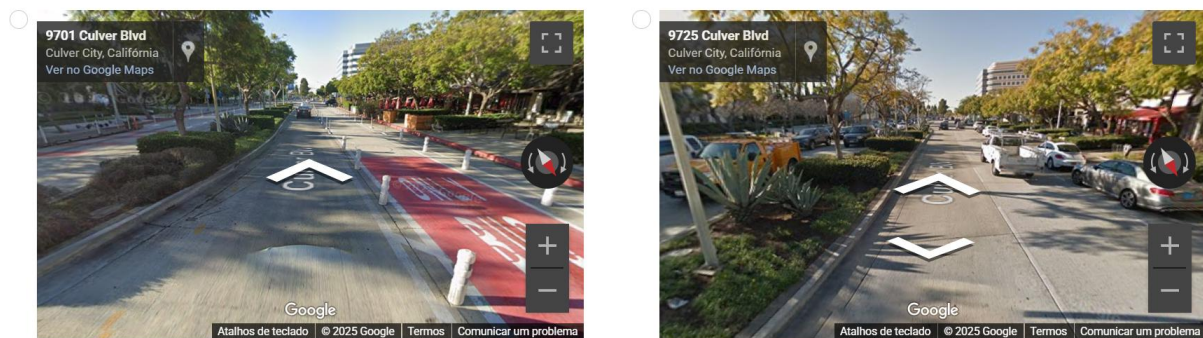


Figure 2 - Example before and after image pair shown to respondents. ([Lopes, 2025](#))

For every pair shown, participants were asked to indicate their preference by choosing either the before or the after street scene (without knowing the images' dates). Alongside this binary preference measure, respondents evaluated the impact of seven perception items on their choice rated on a five-point Likert scale (shown in Figure 3): comfort, liveability order, safety, security, stress and usefulness. In addition, a contextual question distinguished respondents with direct familiarity with the location.

*Evaluate to what extent the following factors have **influenced** your decision.

	Strong Negative Influence	Negative Influence	No Influence	Positive Influence	Strong Positive Influence
Usefulness (e.g., commuting, access to restaurants, parking, socializing, etc)	☐	☐	☐	☐	☐
Safety (e.g., traffic accidents)	☐	☐	☐	☐	☐
Security (e.g., crime)	☐	☐	☐	☐	☐
Stress (e.g., traffic)	☐	☐	☐	☐	☐
Comfort (e.g., shade, seating, etc)	☐	☐	☐	☐	☐
Order (e.g., structured public space)	☐	☐	☐	☐	☐
Liveability (e.g., recreational/green spaces, sense of community, etc)	☐	☐	☐	☐	☐

Figure 3 – Likert scale evaluation of influence of the seven emotions. ([Lopes, 2025](#))

The final section of the survey, collected sociodemographic information, including gender, age, nationality, occupation, commuting modes used, average commute time and the number of distinct commuting modes typically used. Further questions asked about household characteristics such as car ownership, the presence of children, pet ownership (dogs and other pets), whether they had specialist knowledge in urban planning, their place of residence (which enabled deriving their migrant status) and

the availability of parking at both home and work.

From the wide set of variables collected, only a subset was integrated into the present dissertation, which can be seen in Table 1.

Table 1 – Survey variables used and respective variable types.

● preference	● occupation	● age
● context	● urban_specialist	● gender
● has_car	● has_child	● comfort
● parking_home	● pet_dog	● liveability
● parking_work	● pet_not_dog	● order
● main_commute_type	● nationality_grp	● safety
● commute_mode_count	● residence_grp	● security
● commute_time	● is_migrant	● stress
		● usefulness

● Binary ● Categorical ● Discrete

The main outcome was the binary preference variable, while the seven perception items were used to construct latent factors in later stages. The contextual indicator was retained to account for differences in familiarity across respondents and with the “Not In My Back Yard” (NIMBY) effect, as well as sociodemographic characteristics included as moderators within the modelling framework. Other variables from the original questionnaire were not employed, either because they overlapped with the selected set, displayed insufficient variability or frequency in category observations or did not align with the analytical framework of this dissertation.

An additional consideration was that the original survey evaluated eight interventions per respondent. This aspect was particularly relevant for the modelling stage of the research, as not all interventions were retained in the final dataset due to missing or unsuitable images. It also bases the structure of the predictive prototype, where intervention identifiers are used to define holdout groups to guarantee strict separation between training and tuning samples.

3.3 Image collection and preprocessing

The dataset of before and after street images was assembled entirely from open or self-produced sources, thereby avoiding copyright restrictions, associated with using offline versions taken from Google Street View, that apply to all academic, nonprofit and commercial projects (Figure 4).

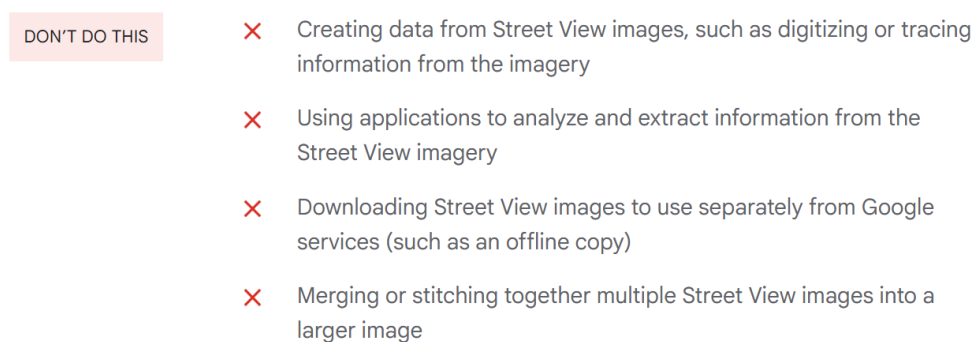


Figure 4 – Google Street Views copyright restrictions. (Google, 2025)

Image pairs were primarily collected from crowdsourced street-level databases including Mapillary

and KartaView, supplemented by official project websites, architectural publications and municipal news articles that documented street transformation projects. In the case of Lisbon's interventions, when no suitable images could be located, photographs were taken directly. In a limited number of cases, manual edits were performed when only one view of an intervention was available: before images were digitally altered to approximate the after condition and vice versa. A graphical summary of the images' source types can be analysed in Figure 5 and a more in-depth reference summary can be found on the GitHub repository, where all scripts and results are published, (Alves, 2025).

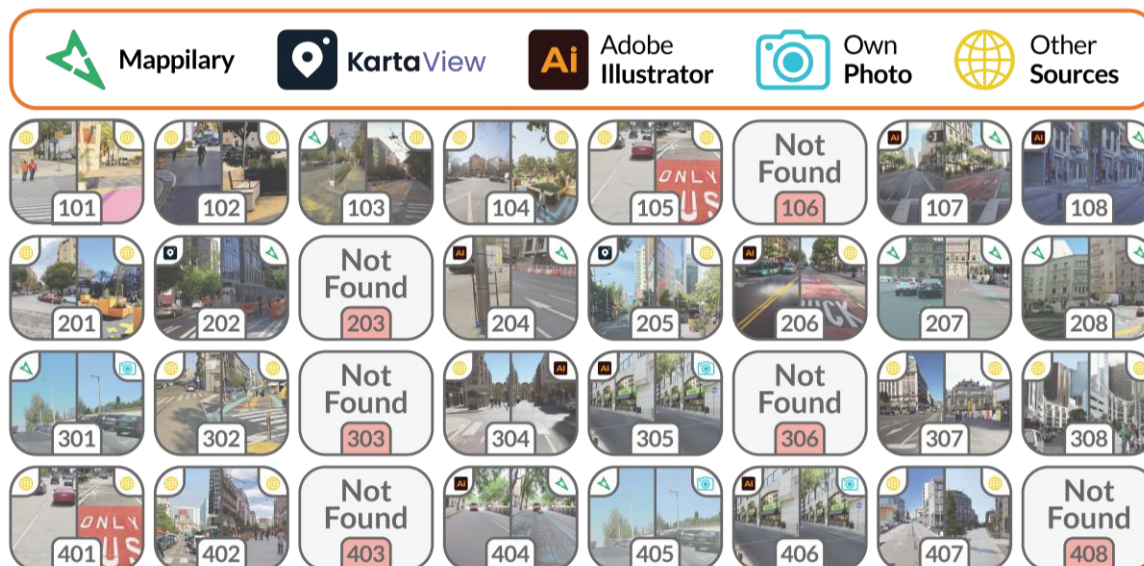


Figure 5 – Graphical summary of the image sources used per intervention.

Each pair represents a real-world intervention and constitutes the unit of analysis for the subsequent colour and segmentation processing. Thus, to ensure comparability, preprocessing included cropping and recentring to align the framing of each before and after view as closely as possible. This reduced the influence of perspective differences and ensured that subsequent measures of visual change reflected the intervention rather than artefacts of misalignment.

The original survey was designed around 28 interventions, spanning car-focused redesigns, transit improvements, pedestrianisation and cycling infrastructure projects. Respondents were presented with two interventions of each type, grouped into four pre-defined blocks of image pairs within the LimeSurvey platform. Due to this grouping structure, some interventions appeared in more than one block, increasing their sample size of observations.

From the original 28 interventions, two car-oriented, two pedestrian-oriented and one transit-oriented interventions had to be excluded due to the absence of appropriate before and after photographs, as is shown in Figure 5. Specifically, this corresponded to one missing intervention from groups 1 and 2 and two interventions each from groups 3 and 4. Therefore, the final dataset retained 23 out of the 28 unique interventions.

For several of these transformations, more than one candidate pair of images was found, often differing in angle orientation, lighting and weather conditions, time since intervention or image quality. To avoid duplicating individual survey observations and to maintain consistency across the dataset, only a single pair of images was retained for each intervention. The choice was made manually, prioritising

the best available quality and the pair that offered the closest alignment between before and after views.

Allowing multiple pairs for the same intervention would have artificially inflated the number of observations, repeated respondents' answers and over-represented certain projects. This is analogous to the problem of double-counting in evidence synthesis, which can distort effect estimates and overstate precision (Hussein et al., 2022). This issue would have been particularly problematic since some interventions were richly documented with several available photographs, while others were not. By standardising at one image pair per intervention, the analysis preserved the impartiality of the survey's structure and ensured that each intervention contributed equally to the subsequent modelling.

3.4 Colour data gathering

An essential stage of the workflow involved the systematic quantification of colour-related characteristics in the before and after street images, since visual appearance is among the cues through which people evaluate changes to the built environment (Jaglarz, 2023). The colour analysis served two principal aims: first, to measure the magnitude of perceptible colour change between the paired photographs; and second, to describe the compositional and perceptual qualities of each scene, such as the proportion of green-coloured pixels or the vividness of the overall palette. The values were later related to perceptions and preference.

A reproducible Python workflow was implemented to extract pixel-level descriptors, aggregate them into interpretable image-level statistics and provide visual diagnostics for quality control. The process was divided into four complementary components: general colour change (also known as colour difference or distance), hue presence, brightness and saturation, therefore encompassing all image comparison descriptors.

To ensure pixel-wise comparability, all paired images were first resized to identical dimensions and converted from OpenCV's BGR format to standard RGB. This preprocessing step was necessary for subsequent operations such as colour differencing and to produce aligned diagnostic outputs.

Overall colour change was calculated in the CIELab* colour space, which is designed to approximate perceptual uniformity (Luo et al., 2001). A raw Euclidean distance between two colours already provides a measure of difference, but it does not account for known non-uniformities in human colour perception.

For example, the raw Euclidean distance for a pixel that changed from vegetation to asphalt would be 51 units; the ΔE_{2000} correction reduces this to around 48 units, remaining above the threshold of 30, used in this methodology, signifying that the change is perceptible. By contrast, in a pixel variation between vegetation in sunlight and the same vegetation in shadow, the ΔE_{2000} value would fall to approximately 3.5 units, which lies below the perceptibility threshold. This demonstrates how the metric distinguishes between meaningful chromatic changes and minor variations due to illumination.

The ΔE_{2000} formula (Equation 1) was therefore applied, refining the Euclidean distance with correction terms and weighting functions for lightness, chroma and hue. This formula is recognized as corresponding closer to the way human observers perceive small colour differences (Habekost, 2013). These refinements impact results particularly in regions where small hue shifts can be disproportionately salient or where chroma effects dominate.

$$\Delta E_{2000} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{k_C S_C}\right) \left(\frac{\Delta H'}{k_H S_H}\right)}$$

Equation 1 - CIE ΔE_{2000} .

In Equation 1, $\Delta L'$ is the difference in lightness, $\Delta C'$ is the difference in chroma and $\Delta H'$ is the difference in hue. The terms S_L , S_C and S_H are weighting functions that adjust for the non-uniform sensitivity of human vision. The parameters k_L , k_C and k_H are correction factors that can be set according to experimental conditions but are equal to one under standard conditions. The final term, R_T , is a rotation factor constant that accounts for interactions between hue and chroma differences in specific regions of colour space, most notably the blue region.

From the pixel-wise ΔE_{2000} , three scalar variables were derived: the mean ΔE_{2000} , representing the average perceptual change across the image; the standard deviation of ΔE_{2000} , indicating whether change was evenly distributed or concentrated in certain areas; and the percentage of pixels with ΔE_{2000} values greater than thirty.

While ΔE_{2000} thresholds reported in lab material studies usually fall between one and five, these reflect controlled viewing and are therefore not directly applicable. To ensure robustness to lighting and viewpoint noise in the gathered street-view images, the operating threshold was derived from empirical data which served as a pragmatic result for clearly noticeable change. A comparison between using the standard values and the operating threshold chosen can be seen in Figure 6.

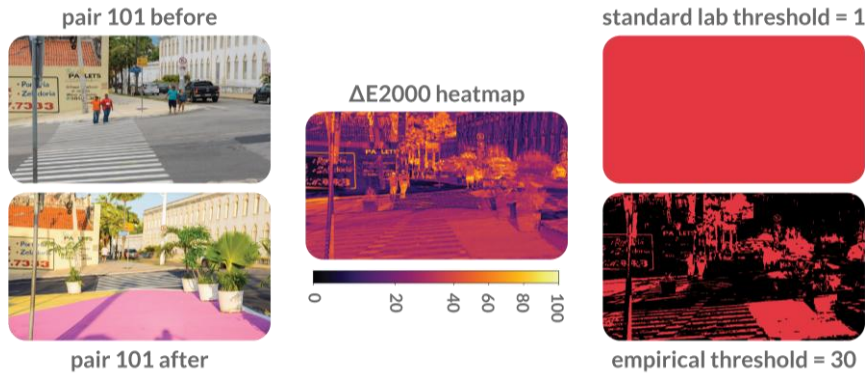


Figure 6 – Comparison between standard lab and empirically chosen thresholds.

While only the threshold variable was retained as a variable for subsequent analysis, to avoid collinearity, the other measurements, heatmaps and binary masks of ΔE_{2000} were produced to allow for quality control.

The colour palette of each image was then examined in HSV (Hue-Saturation-Value) space, which separates chromatic content (hue) from perceptual attributes (saturation and brightness). Each pixel was first binned into one of six hue ranges (red orange, yellow, green, blue or purple) with applied thresholds on saturation and value, which prevented noise from greys or very dark pixels that carry little chromatic information. The thresholds and bins used can be analysed on the GitHub repository (Alves, 2025) and are illustrated in Figure 7.

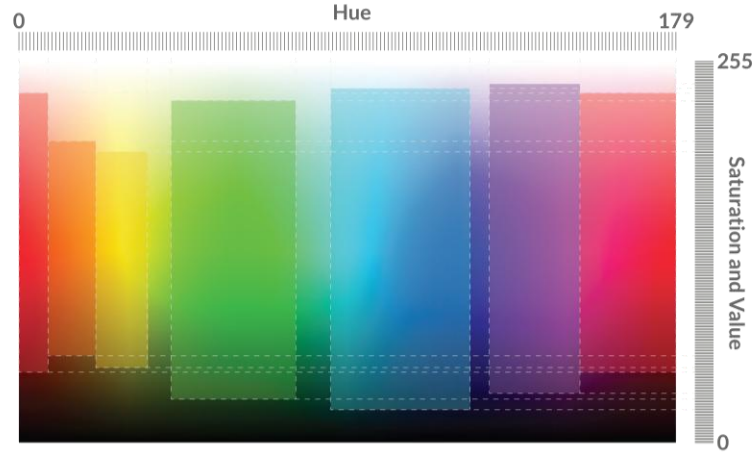


Figure 7 – Representation of the colour presence bins and thresholds applied.

Hue binning is a common way to analyse colour presence and can be found in Bhunia’s further research into a novel feature descriptor for image retrieval (Kumar Bhunia et al., n.d.) or in “*Entropic Analysis of Colour Homogeneity*” (Alemaskin et al., 2005) which closer represents the methodology used in this dissertation.

For each image pair, the variation in the proportion of each category was computed. To summarise the overall diversity of the palette, the Shannon entropy (Equation 2) was calculated across the hue proportions, following the workflow defined by Alemaskin (2005):

$$H = - \sum_{i=1}^6 p_i \ln(p_i)$$

Equation 2 - Shannon entropy.

, where p_i is the proportion of pixels belonging to hue i . This value was then normalised to the interval zero to one by dividing by the maximum possible entropy $\ln(6)$, which corresponds to a perfectly even distribution of colours. Thus, higher entropy values indicate greater colour diversity, while lower values signal dominance by one or two hues.

This entropy measure captures not only the presence of multiple colours, but their relative balance and index of palette complexity. In other words, an image in which blue accounts for 42% of pixels, while the remaining 3.2% detected are split among the other five hues, will have an entropy value around 0.15, reflecting strong dominance by a single colour and limited diversity. By contrast, if all six hues are closely represented, the entropy can reach higher values, reflecting balance and variety in colour composition, as depicted in Figure 8.

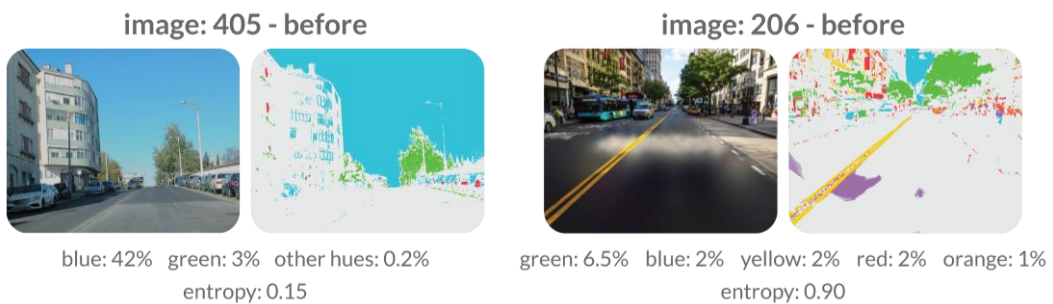


Figure 8 – Shannon entropy: interpretability example.

Brightness and saturation were then analysed using the V and S channels of HSV, respectively. The mean brightness was computed for each image to index overall lightness, while the standard deviation of brightness measured heterogeneity in illumination. The mean saturation captured how vivid the scene appeared overall and the standard deviation of saturation indicated the degree of variability in colour intensity across the image. For both brightness and saturation, the interquartile range, heatmaps and data summary plots were also computed to allow for quality control. Examples of these data summary plots for brightness and saturation results can be analysed in Figure 9 and Figure 10, respectively.

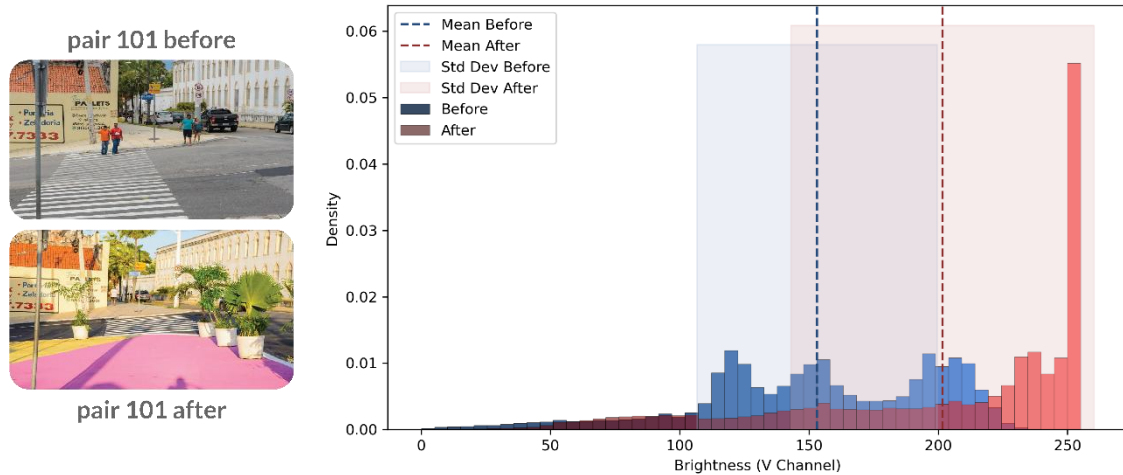


Figure 9 – Intervention pair 101 and respective brightness data summary plot.

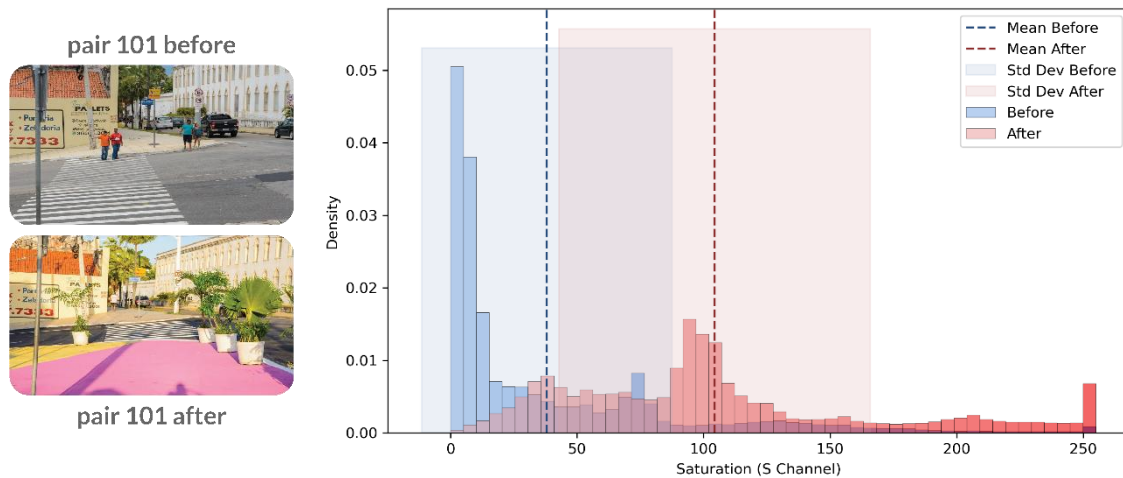


Figure 10 – Intervention pair 101 and respective saturation data summary plot.

The final set of variables obtained through this workflow is summarized in Table 2.

Table 2 – Colour variables used and respective variable types.

● blue_var	● orange_var	● red_var
● green_var	● purple_var	● yellow_var
● color_entropy_var	● mean_brightness_var	● std_dev_brightness_var
● prctdeltaE_above_30	● mean_saturation_var	● std_dev_saturation_var

● Continuous

It consists of the percentage of pixels above the ΔE_{2000} threshold of thirty (clearly noticeable changes), the variation in proportions of red orange, yellow, green, blue and purple pixels, the Shannon

entropy of hue proportions (palette diversity), the mean brightness, the standard deviation of brightness, the mean saturation and the standard deviation of saturation.

Together, these twelve variables provide a compact yet comprehensive set of descriptors that are perceptually meaningful, easily interpretable and suitable for statistical modelling.

3.5 Segmentation model development

Pixel-wise semantic segmentation is a computer vision technique that assigns a class label to every single pixel in an image. In the context of this dissertation, segmentation allows raw street images to be converted into measurable data with each pixel being linked to a meaningful class, making it possible to calculate how much of the image area is covered by road, vegetation or other elements before and after an intervention. A pixel-wise segmentation example is illustrated in Figure 11. The results from this process were added to the foundation of the later stages of analysis, where physical changes in the street environment can be quantified and statistically related to public perceptions and preference over urban interventions.



Figure 11 – Representation of a pixel-wise segmentation step.

To predict segmentation masks for every intervention image collected, a DeepLabV3 model with a ResNet-101 backbone was selected and fine-tuned on the Cityscapes dataset (Cordts et al., 2016). This final composition passed through several iterations in both the base model and backbone, as well as in the fine-tuning dataset and strategy.

DeepLabV3 is a fully convolutional network-based convolutional neural network (FCN-based CNN) architecture for semantic segmentation, enhanced with atrous convolutions and atrous spatial pyramid pooling (ASPP). It first introduced the ASPP system, which allows the model to look at image features at multiple scales simultaneously (L.-C. Chen, Papandreou, Schroff, et al., 2017). This is particularly important in street scenes, where objects can vary dramatically in size, from large continuous surfaces such as roads or buildings to small and thin features such as traffic signs, bicycles or leaves. By aggregating information across different spatial scales, the model achieves higher accuracy in detecting both large and small elements within the same image.

The backbone used inside DeepLabV3, was ResNet-101 and is itself a deep convolutional neural network (DCNN) originally developed for large-scale image recognition tasks. The key innovation in

ResNet architectures is the use of residual connections, which allow very deep networks to be trained without suffering from the vanishing gradient problems (K. He et al., 2015). In practice, this means the model can learn both simple, low-level patterns (such as edges, corners or textures) and more complex, high-level representations (such as the general outline of a vehicle or the canopy of a tree). Thus, the depth of ResNet-101 provides sufficient capacity to capture the wide diversity of features present in urban street imagery (L.-C. Chen et al., 2018).

Alternative backbone networks were evaluated before finalising the choice of ResNet-101. Lighter backbones, such as ResNet-18 or ResNet-34, were computationally efficient but produced noticeably weaker segmentation performance, particularly in detecting smaller or less common classes. On the other hand, heavier backbones such as ResNeXt-152 variants were found to be excessively resource-intensive, requiring higher-end graphics processing units (GPUs) and memory capacities that would make the workflow impractical for replication on standard hardware. ResNet-101, therefore represented the optimal trade-off, by providing strong segmentation accuracy across all relevant classes while remaining feasible to train and run on a single regular-performance workstation.

Finally, DeepLabV3 itself was chosen over other base architectures, such as U-Net or SegNet, because it is better suited for the complexity of urban imagery (Gurita & Mocanu, 2021). U-Net, while effective in medical imaging and smaller-scale problems, tends to struggle with the diversity of objects present in street scenes (Ronneberger et al., 2015). SegNet, though efficient, lacks the multi-scale feature extraction capabilities that are crucial for accurate performance in this context (Badrinarayanan et al., 2015).

Additionally, the model was not trained entirely from scratch but instead was initialised with pretrained weights obtained from the COCO dataset, a large and diverse collection of everyday photographs used in computer vision research. These weights represent parameters that have already learned to recognise a wide variety of basic visual features such as edges, textures, shapes and colours. Reusing such pretrained weights is a strategy known as transfer learning. It accelerates the training process because the network does not have to rediscover low-level patterns that are common across many types of images and it also improves generalisation because the pretrained filters already capture visual regularities that are not specific to a single dataset.

For example, the texture of asphalt or the outline of a tree canopy can be learned from COCO and then applied effectively to street scenes. This is standard practice when using DeeplabV3 in Python (L.-C. Chen, Papandreou, Schroff, et al., 2017). Only the final classifier layer of the network had to be replaced, because the workflow proposed required predictions across nineteen specific classes rather than the eighty object categories defined in COCO or the thirty-four class IDs available when using Cityscapes.

As mentioned, the source dataset used in training, Cityscapes originally defines thirty-four distinct semantic labels. However, many of these are either semantically redundant (for example, different types of road markings) or not relevant for analysing perceptual responses to urban interventions (such as the ego-vehicle, which only appears in the dataset because the images were captured from a car). To address this, the labels were remapped into nineteen training IDs following the standard Cityscapes mapping scheme (Cityscapes Dataset, 2025).

Consolidating the labels not only reduces model confusion between categories that are visually similar but also ensures that the network concentrates its learning capacity on classes that meaningfully

contribute to the analysis of urban interventions. Labels that were irrelevant or too specific, such as license plates or reflections, were therefore reassigned to the ignore index (255). This label ensured that those pixels were excluded from the loss function during training, so the model did not waste capacity trying to predict categories that were of no use for the later perceptual analysis.

To ensure consistency between inputs and their ground-truth labels, both images and their corresponding annotation masks were processed in parallel. Before being passed to the network, images and masks were resized to 256×512 pixels. This decision reflects a compromise between larger resolutions that provide more detail, which helps the detection of small and thin objects such as poles or pedestrians and the imposition of a much heavier burden on GPU memory and computational speed. A resolution of 256×512 was found to preserve sufficient visual information for accurate segmentation while keeping the training process feasible on the available hardware.

Still, real-world urban environments vary in terms of perspective, occlusion and lighting conditions. Hence, a model trained only on unaltered images would risk overfitting to the specific conditions of the training set and could fail to generalise to new streets or photographs taken under different weather or lighting conditions. To avoid this issue, the training pipeline incorporated data augmentation, a technique that artificially increases dataset diversity by applying random transformations to the images. Importantly, these augmentations had to be applied in a synchronised way to both the image and its corresponding mask, so that the labels always remained perfectly aligned with the transformed image. The chosen augmentation steps are exemplified in Figure 12.

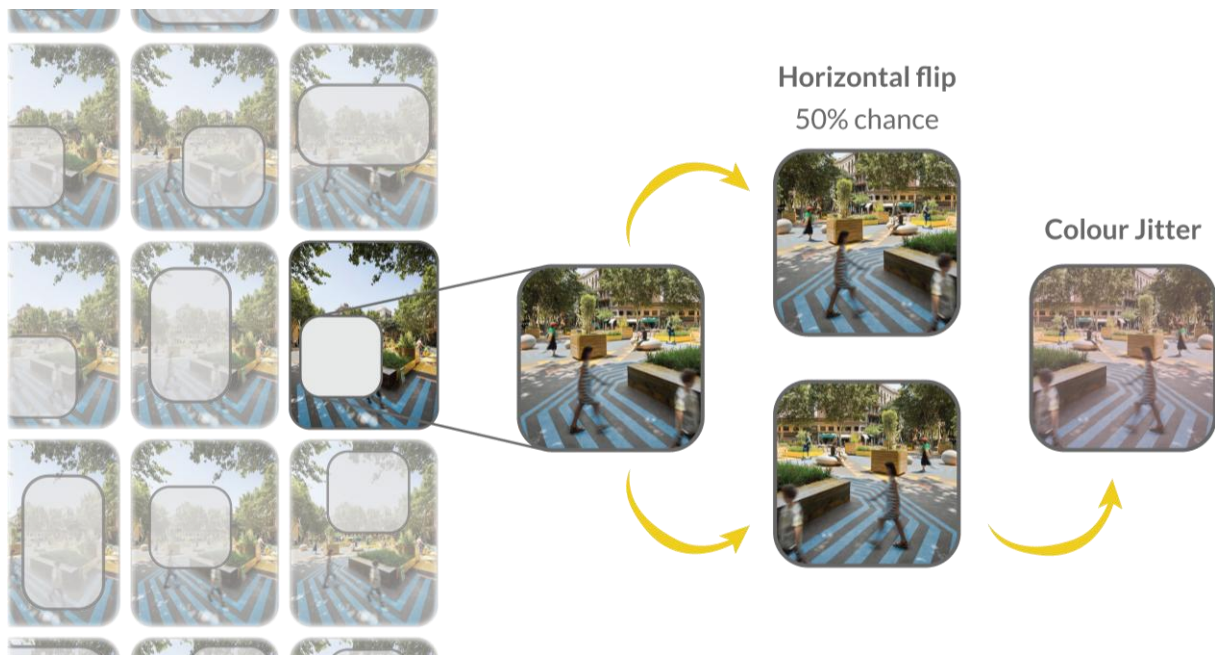


Figure 12 – Representation of the data augmentation steps used.

These included resized crops with a scale factor between 0.5 and 2.0 and aspect ratio between 1:1 and 2:1, which expose the model to partial views and perspective distortions without zooming in so much to the point where no object can be defined; horizontal flips with a 50% probability, which simulate mirrored viewpoints; and colour jitter, which randomly alters brightness, contrast, saturation and hue to replicate variations in weather, time of day or camera settings. These augmentations help the model learn to recognise classes robustly even when appearances change.

After augmentation, all images were normalised using the standard statistics of ImageNet. Normalisation rescales pixel intensities to a standard range, ensuring that the inputs remain numerically stable and compatible with the pretrained weights originally obtained from COCO.

Training the neural network itself involved teaching it how to compare its predictions with the correct answers and then gradually adjusting its internal parameters to reduce the difference. In this dissertation, that comparison was conducted using a function called weighted cross-entropy (CE) loss. A loss function is a mathematical tool that measures how wrong the model's predictions are. CE loss is widely used in classification tasks because it penalises incorrect predictions more heavily when the model is confident but wrong (i.e., when the model assigns large probability to an incorrect class) and it rewards high confidence in the correct class (Lightly, 2025).

The weighted version used allows certain classes to count more in this calculation than others. This was necessary because in urban street images, some elements appear much more often than others. For example, pixels belonging to the sky or buildings are abundant, while pixels representing bicycles or small signs are rare. If no adjustments were made, the model would appear inaccurate by focusing on predicting uncommon classes as much as common ones, even if it consistently failed to identify roads, buildings or the sky. To correct this imbalance, weights from 4 to 5 were assigned to road, sidewalk and vegetation, because these are the elements most often present in a street image. Other classes, such as fences or poles, received smaller weights between 1 and 3, depending on their frequency on the dataset. This way, mistakes on the most common elements were penalised, forcing the model to learn them well.

Once the loss function defined how wrong the model was, an algorithm was needed to update the network's parameters so that the loss decreases over time. In this case, the AdamW optimiser was used. An optimiser is a set of rules that decide how much to change each parameter in response to the loss. AdamW is an improved version of the Adam algorithm, which adapts the learning rate for each parameter individually depending on how large or small its gradients are (Loshchilov & Hutter, 2019). A gradient represents the direction and magnitude of change that should be applied to each parameter to reduce the overall error and is computed as the derivative of the loss with respect to a given parameter. A positive gradient means that increasing the parameter would increase the error, while a negative gradient means that increasing it would reduce the error.

The W in AdamW stands for weight decay, which is a form of regularisation that discourages the parameters from becoming excessively large (Loshchilov & Hutter, 2019). Exceptionally large weights can lead to overfitting, where the network memorises the training images instead of learning general patterns. By including weight decay, the model is nudged toward simpler and more general solutions.

This optimiser was configured with an initial learning rate of 0.0006, which controls the overall size of each update step and a weight decay coefficient of 1×10^{-4} , which acts as regularization by penalising large weights and controlling model complexity (Dive into Deep Learning, 2025; van Laarhoven, 2017). These values are the results of an empirical tuning process: the learning rate is large enough to make meaningful progress at each step, but not so large that the model would diverge; the weight decay is strong enough to reduce overfitting, but not so strong that it prevented the network from learning detailed patterns. This kind of hyperparameter trade-off is widely acknowledged in deep learning as a "*black art*"

that must be tuned via experimentation (Smith, 2018).

Another challenge in training large networks is memory capacity. Each update requires storing information about the gradients for all parameters and, on a typical GPU, it is not possible to fit many images in memory at once. To address this, gradient accumulation was used. Instead of updating the network after every small batch of images, the model processed four batches in sequence, calculated their gradients and added them together. Only after the fourth batch were the parameters updated.

From the optimiser's point of view, this is almost the same as if all those images had been processed at once, so it simulates a larger batch size without requiring more memory at any given moment (Mindspore, 2025). In this setup, the batch size was set to five images so, after accumulating gradients across four batches, the effective batch size seen by the optimiser was twenty.

To further improve efficiency, mixed-precision training was enabled. This means that some calculations were performed using 16-bit floating-point numbers instead of the usual 32-bit precision. Using fewer bits reduces memory usage and allows faster computation, but it also increases the risk of numerical errors when numbers are exceedingly small. To prevent this, a gradient scaler automatically adjusted the values during training, ensuring that they stayed within a safe range. Gradient norms were also monitored, since gradients can become unstable due to a problem known as exploding gradients, forcing the model to fail to learn. In combination, gradient accumulation and mixed precision allowed the model to be trained effectively on limited hardware while still benefiting from the stability of larger batch sizes (Business Analytics Institute, 2025).

Lastly, the scheduler by which the learning rate changes during training plays a critical role in whether the model converges to a good solution (Smith & Topin, 2018). If the learning rate stays constant, training may either stall (if it is too low) or diverge (if it is too high). To avoid this, a dynamic learning rate rule called OneCycleLR was applied. This scheduler increases the learning rate gradually from its initial value up to a peak of 0.0015 and then decreases it smoothly back down toward a small value. The idea behind this one-cycle policy is to allow rapid exploration of the parameter space early on (when the learning rate is high) and then fine-tuning around a good solution later (when the learning rate is low) (Smith & Topin, 2018). This pattern can be observed in the evolution of the model's capacity, depicted in Figure 13.

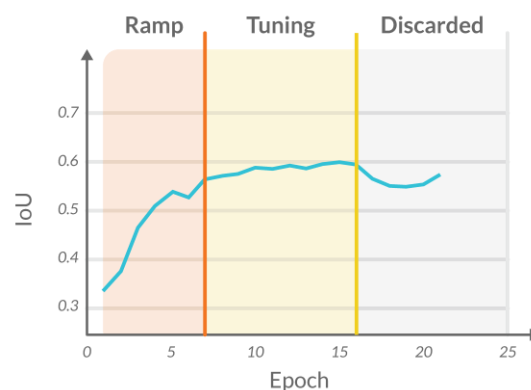


Figure 13 – Segmentation model training evolution.

Specific parameters controlled the exact shape of the cycle. The `pct_start` parameter was set to 0.3, which means that the first 30% of the cycle was spent increasing the learning rate. The `div_factor` was

25, which set the initial learning rate to 1/25 of the maximum, ensuring a gentle start. The `final_div_factor` was 100, which defined the ratio between the maximum and the final minimum learning rate. Once again, these parameters reflect a thorough tuning process to achieve the best loss function results. Together, these settings encouraged the model to learn quickly at the beginning and to stabilise at the end, a strategy that has been shown to improve both the speed of convergence and the quality of the final model (Smith & Topin, 2018).

To train the model, 2975 images were used from the Cityscapes `leftImg8bit` dataset with their corresponding `gtFine` masks, which is a large collection of vehicle point of view (POV) photographs from central European city streets, whose locations are shown in Figure 14. Cityscapes provides paired images from two dash cameras on a vehicle (left and right), but only the left-camera images were used. This choice made the setup simpler, prevented duplicate-looking data and still gave a wide variety of street scenes to train on. The choice between the right or left camera was arbitrary.



Figure 14 - Map of `leftImg8bit` locations from the Cityscapes Dataset. (Cityscapes Dataset, 2025)

As preparation for training and to make sure that the exact same batches of images could be used again in future experiments, the list of chosen IDs per batch was saved. The selection process itself was random but controlled with a seed (42). By setting the seed to 42, every time the training is run, the computer will choose the same random order.

For the validation set (the images used to check how well the model is learning), randomness was avoided altogether and instead, the same fixed order was used. This made sure that the validation results were perfectly consistent and comparable across different model setups when tuning the previously referred parameters.

Training models require moving around large amounts of data, which can be done using worker processes. Workers are helper programs that run in parallel to speed up data loading. In theory, using more workers makes data preparation faster (PyTorch - `torch.utils.data`, 2025), but in practice, it can sometimes cause system errors or crashes depending on the environment used. To avoid these issues, no extra workers were added, meaning all data loading was managed by the main process. This was a slower process but much safer for reproducibility.

Another key issue was memory management since the GPU has a limited amount of memory available. If this memory fills up, training will crash. To prevent it, unused data was cleaned up after

each training step (a process called garbage collection) and the GPU's memory cache was regularly cleared.

During training, progress was tracked using a system of checkpoints. A checkpoint would save the current state of training so it could be resumed later without starting over, in case of a system shutdown. These contained the model state dictionary (all the learned parameters of the network), the optimizer state (information about how the learning algorithm was updating those parameters), the current position in training (epoch number and batch index), the history of losses (evolution seen so far), the best validation scores achieved and additional tracking measures such as the cumulative Intersection over Union (IoU).

Alongside checkpoints, the training loop also wrote detailed log files. These logs recorded, for every batch, how long processing took, how much GPU memory was used, how much central processing unit random access memory (CPU RAM) was used, the learning rate at that moment, the raw loss values and the IoU score.

For visual diagnostics, the predicted segmentation masks (which are arrays of numbers) were converted into colour images using a custom palette, where each class (e.g. road, sidewalk, vegetation) was assigned to a unique colour. A small subset of validation predictions was saved this way after each epoch, so that mid-run quality checks of the model setup could be feasible without interrupting it.

After each epoch, the model's performance was evaluated using the previously referred Intersection over Union (IoU) metric, which is widely used for segmentation tasks ([Garcia-Garcia et al., 2017](#)). IoU measures how well the predicted mask overlaps with the ground truth (the correct handmade labels). It is computed by dividing the area of overlap by the area of union between the two shapes. The value ranges from 0 (no overlap at all) to 1 (perfect agreement). This step is illustrated in Figure 15.



Figure 15 – Intersection over Union representation.

To get a broader sense of accuracy, two types of IoU results were tracked. First, class-wise IoU was computed, which gave one IoU value for each semantic category (e.g. road, building, sky). This helped identify weaknesses in specific areas. For instance, the model might perform well on “road space” but poorly on “bicycles”. Second, the mean IoU was calculated by averaging across all classes that had at least one pixel present in the batch. This gave a single number summarising overall accuracy.

For the model's decision-making, a special third metric called mean relevant IoU was also defined. This was the average IoU across a smaller group of classes (road, sidewalk, vegetation and sky), since these are the most frequently present classes in street images and therefore the most important to predict right, so later correction stages are more efficient.

Training was scheduled to last for up to 100 epochs (one epoch equals a complete pass through all the training batches). However, to avoid time wasting, speed up model setup iteration and to reduce the risk of overfitting (when a model memorises the training data instead of learning general patterns), an early stopping rule was used. Specifically, if the mean relevant IoU failed to improve for ten consecutive

epochs, training was halted. This patience value of ten struck an empirical balance between giving the model enough time to improve slowly but prevented endless training once learning had clearly plateaued.

Whenever the checkpointing system detected that the mean relevant IoU had improved, a new best-performing version of the model was saved and patience was set back to zero. This confirmed that the model was gradually learning to represent the categories most important for perceptual analysis. A comparison chart between the top five model setups trained can be analysed in Figure 16.

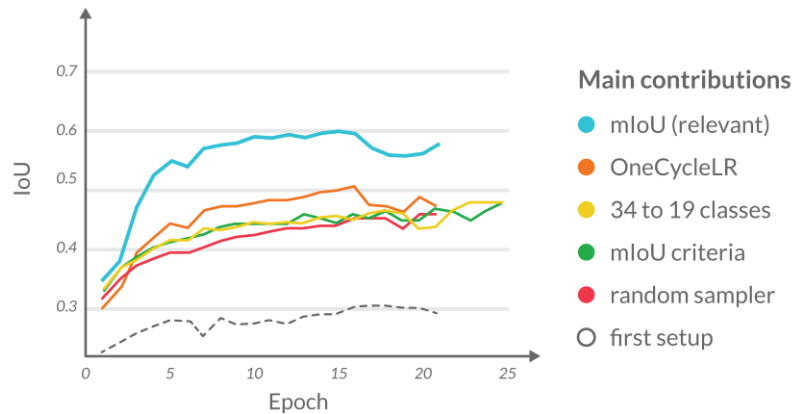


Figure 16 - Semantic segmentation model's setup comparison.

The noticeable jump in improvement from the second-best model to the best model in Figure 16, is due to the change in validation criteria, described previously as the mean relevant IoU function, which focused model training into the most frequent classes, enabling higher fidelity results.

When training ended, the system kept the best validation checkpoint encountered during the run. This checkpoint contained the model's learned weights along with all progress metadata, making it possible to reload the same state later. That way, inference (making predictions on new images) could be performed with the most successful version of the model, not just the one from the final epoch.

3.6 Segmentation inference and variation measurement

Once the model had been trained and validated, the next step was to apply it to the gathered before and after intervention images selected for analysis. This stage is referred to as inference, meaning taking a trained model and using it to make predictions on new, unseen data.

Each input photograph was resized to the same dimensions used during training to ensure compatibility, passed through the best training model and converted into a predicted segmentation mask, where every pixel had one of the nineteen class labels used in the Cityscapes training scheme, such as road, building, vegetation, sky or person. The predictions were then rendered as coloured images using the same custom palette as in training.

This colouring step makes the outputs legible for visual inspection, since each class is represented by a distinctive colour that remains consistent across all images. The predictions were also resized back to their original resolution, so that the segmentation masks could be directly compared to the source photographs without losing detail.

At this stage, a major limitation of automatic segmentation had to be addressed. Although the trained network was effective at recognising the broad structure of urban scenes, its performance was not

equally robust across all categories. Small or visually variable classes, such as bicycles, bus lanes or cycle lanes, were often missed or misclassified. This is not a flaw of the implementation but rather reflects the training data itself: the Cityscapes dataset, while highly detailed, was developed for central European cities and does not include certain street features that are common in other contexts. For example, dedicated cycle lanes and bus lanes vary widely across cities in colour, design and signage, which makes it difficult to include them as a consistent class in standard datasets. For this reason, they were not part of the nineteen categories predicted by the model and therefore could not be produced automatically.

To overcome this issue, a manual correction stage was introduced between the raw inference and the quantitative analysis, following common practice (Reza et al., n.d.). In this step, the coloured segmentation masks were reviewed and adjusted wherever the model had produced obvious errors or where non-standard but analytically important features needed to be added. For example, if the model misclassified a large section of a cycle lane as part of the road or if a bus lane was absent entirely, the mask was manually edited in Adobe Illustrator to assign the correct labels. Similarly, occasional mistakes, such as confusing dense in shadow vegetation with building facades or failing to detect narrow sidewalks, were corrected. This manual stage ensured that the final masks produced a dataset that was accurate enough for subsequent quantitative analysis. Importantly, this step was kept focused on correcting only major and systematic errors, rather than pixel-level noise, so that the process remained efficient and consistent. An example of these corrections can be seen in Figure 17.

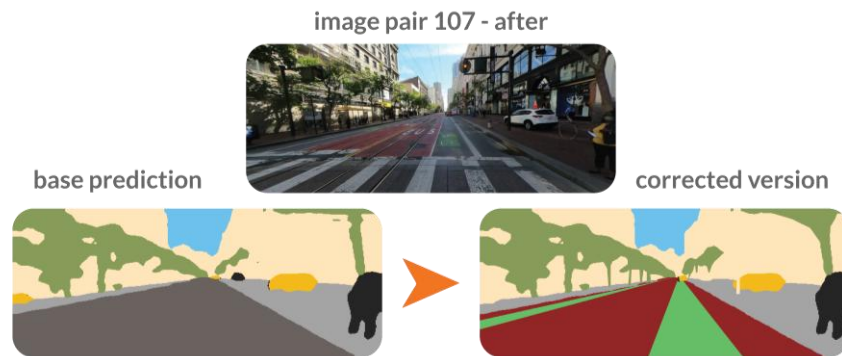


Figure 17 – Semantic segmentation's correction stage example.

Once the corrected masks were available, the next stage was to compare the before and after pairs of images in a systematic way. Each pair was analysed pixel by pixel to determine the distribution of classes within the scene. For every image, the number of pixels assigned to each class was counted and converted into a percentage of the total. This produced a profile of how much of the image area was occupied by elements such as road, sidewalk, building, vegetation, sky, person, vehicle, bicycle, cycle lane or bus lane. By calculating the difference between the before and after percentages, it was possible to quantify the changes that occurred between the two states of the same street. For example, if the proportion of road pixels decreased while the proportion of vegetation and sidewalk pixels increased, this indicated a street transformation that shifted space away from car-oriented urbanism and towards greener pedestrian areas.

The comparison results were exported in two forms. First, the pixel proportions and their differences were stored in a structured comma-separated values file (.CSV), providing a numerical dataset suitable

for statistical analysis and integration with the perception models. Second, the differences were visualised as bar plots for each image pair, showing which classes had grown or shrunk in relative size. These plots served as an immediate quality control feature and provided intuitive illustrations of the transformations captured by the images. An example of this representation is shown in Figure 18.

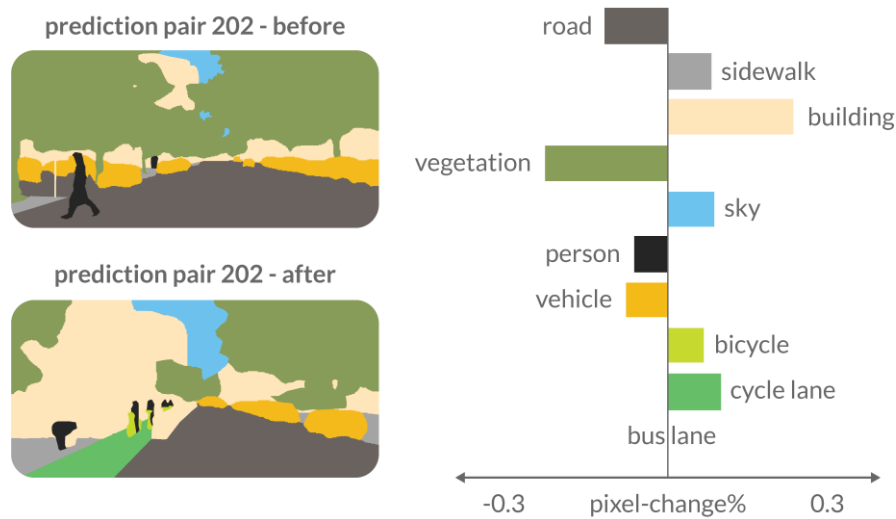


Figure 18 – Example of a percentage of pixel-class change plot.

3.7 Preliminary statistical analysis

The survey, colour and segmentation datasets were first examined through a set of preliminary analyses to describe the variables, assess their relationships and investigate the structure of the perception measures, as suggested in “*Principles and Practice of Structural Equation Modelling*” (Kline, 2016). The analyses were organised into univariate, bivariate and factor analyses components, providing both a detailed view of individual variables and a sense of how they relate to underlying constructs. These steps established an empirical basis, complementing the theoretical framework and ensured that the predictors included in the structural models were statistically robust and conceptually well defined.

3.7.1 Univariate analysis

The univariate analysis step was defined to provide an overview of the data before moving into more complex bivariate and multivariate modelling and followed several guides on how univariate should be applied to different types of data (Canova et al., 2017; Chen et al., 2024; Marshall et al., n.d.). By examining variables individually, it became possible to characterise their distributions, assess data quality and identify any irregularities that might affect subsequent analyses. Since the data used in this dissertation included several types of variables, different univariate approaches were followed, depending on which type or types were being characterised.

For the continuous variables, such as variation in road space (road_spc_var) or colour differences (deltaE_mean), a set of descriptive statistics was computed. Measures of central tendency included the mode, mean and median, while the spread of the data was defined by the standard deviation, minimum and maximum values, range and the interquartile range (IQR). Skewness was also calculated to evaluate the symmetry of the distributions and potential outliers were counted, using Tukey’s rule

presented in Equation 3, to flag cases that could distort statistical modelling.

Lower fence: $Q1 - 1.5 \times IQR$

Upper fence: $Q3 + 1.5 \times IQR$

Equation 3 - Lower and Upper fences for mild outlier detection.

In Equation 3, the first quartile (Q1) corresponds to the 25th percentile, representing the value below which one quarter of the observations fall. The third quartile (Q3) corresponds to the 75th percentile, representing the value below which three quarters of the observations fall. The interval between Q1 and Q3, known as the interquartile range (IQR), captures the central 50% of the data distribution.

To visualise some of the results from this univariate analysis, summary plots were created for each continuous variable. An example of these can be found in Figure 19.

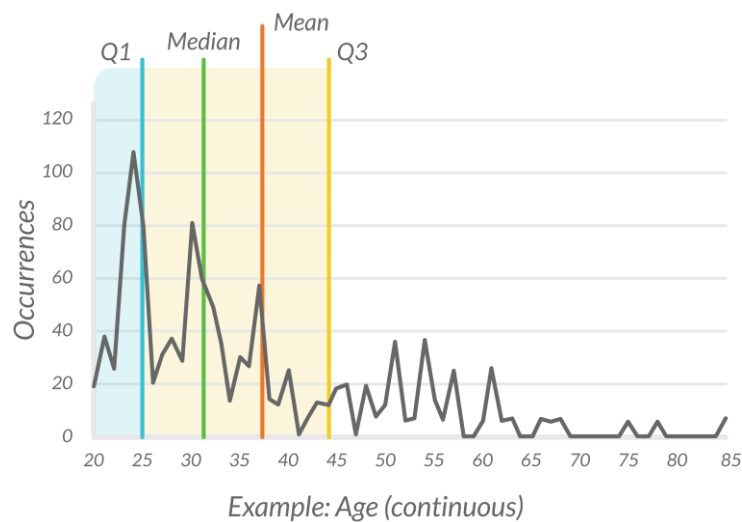


Figure 19 - Example of univariate analysis summary for continuous variables.

The perception variables, measured on a 5-point Likert scale, required a slightly different treatment due to their ordinal nature rather than continuous type. Comfort, safety, liveability order, stress, security and usefulness were summarised using the mode and the median as indicators of central tendency. Dispersion was represented by the minimum, maximum, overall range and the interquartile range. This approach respected the ordered but non-interval nature of the scales, while still highlighting the extent of agreement or divergence among respondents.

Binary variables, such as migrant status, presence of children in the household, whether respondents were urban specialists, owned cars or pets, were summarised through simple percentages of “yes” and “no” frequency in responses. This helped reveal imbalances in the sample. For example, almost eighty per cent of observations preferred the after scenario of an intervention and only thirteen per cent reported knowing the location in question. These distributions not only described the dataset but also indicated where strong asymmetries might influence subsequent modelling.

Lastly, categorical variables with more than two unordered categories were analysed using frequency distributions. Examples included the main mode of commuting, shown in Figure 20, the regional group of residence, the availability of parking at home and the occupation category. These highlighted broad trends in the sample, such as the predominance of motor vehicle commuting (forty per cent), the

overrepresentation of respondents from Southern Europe (around fifty-nine per cent) and the fact that parking access was often either free or entirely unavailable. Summarising the categorical variables in this way provided a clear view of the sample's demographic and contextual structure, ensuring that subsequent statistical procedures were grounded in a well-understood descriptive baseline.



Figure 20 - Univariate analysis of categorical variables (e.g. main commute mode).

3.7.2 Bivariate analysis

The bivariate analysis examined how pairs of variables correlated across the survey, segmentation and colour data. Because the dataset included a mixture of continuous ordinal and categorical variables, no single statistical measure was appropriate for all cases. Instead, the choice of method was adapted to the types of variables being compared.

When both variables were continuous, for instance in image metrics such as brightness and saturation, the strength of their linear association was assessed using Pearson's correlation coefficient. Pearson's r quantifies the degree to which two continuous variables change together along a straight-line relationship, ranging from -1 (a perfect negative correlation) to +1 (a perfect positive correlation), with 0 indicating no linear association. This measure is sensitive to outliers and assumes normally distributed variables, but it remains the most widely used statistic for continuous data. Alternative approaches, such as Spearman's rho, were considered but reserved for cases where ordinal or highly skewed variables were involved ([de Winter et al., 2016](#)).

In practice, the use of Pearson correlations uncovered a high degree of redundancy between four colour-based predictors, particularly in brightness and saturation. Their strong correlation suggested that keeping all four brightness and saturation features separately (mean and standard deviation for each) would inflate multicollinearity in later models. This observation motivated the dimensionality reduction step described in chapter 3.5 Segmentation model development.

For relationships between binary and continuous variables, point-biserial correlations were employed. This statistic is a special case of Pearson's correlation, adapted for situations where one variable can take only two values (for example, car ownership coded as "yes" or "no") and the other is continuous (such as road space variation). Point-biserial correlation measures the extent to which the distribution of the continuous variable differs between the two groups defined by the binary variable. Other approaches, such as group comparison tests (t-tests), were considered, but this framework was preferred because it yields effect sizes on the same scale as Pearson's, making interpretation across analyses more consistent ([National University, 2025](#)).

In practical terms, these results showed, for example, that car ownership was systematically linked to differences in the observed importance of usefulness in preference choices, highlighting how background factors might shape perceptions of street interventions in later models.

The perception variables themselves were measured on a five-point Likert scale. Since Likert responses are ordinal (they indicate order but not necessarily equal spacing between categories), it would have been inappropriate to use Pearson correlations directly as well. Instead, polychoric correlations were estimated, as they assume that ordinal items reflect underlying continuous attitudes (Holgado-Tello et al., 2008). This was essential for determining whether perception items clustered in meaningful ways. The resulting polychoric matrix, presented in Figure 21, was later used as input to the exploratory and confirmatory factor analyses described in chapter 3.7.3 Polychoric factor analysis on perception variables, where they revealed that items such as comfort, safety and liveability loaded strongly onto common latent factors.

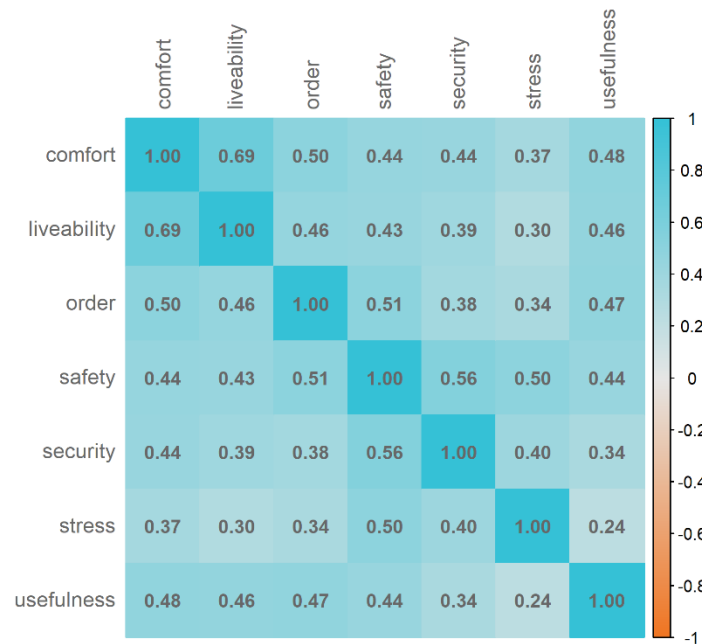


Figure 21 - Polychoric correlation matrix for perception variables.

Without this stage, the latent structure of the perception constructs would have been weaker and less theoretically justified when building the multivariate models.

When both variables were categorical without any natural order, for instance, main commuting mode, region of residence or occupation, Cramér's V was used. This method is derived from the chi-square test of independence and provides a standardised measure of association between nominal variables (University of Regina, n.d.). This identified, for example, strong overlaps between occupation and commuting modes used. It informed later modelling by highlighting whether demographic clustering could influence perceptions of projects when moderating the paths with physical street changes.

For ordinal perceptions compared with continuous colour or segmentation variables, Spearman's rank correlation was used. Unlike Pearson's r , Spearman's ρ does not assume linearity and is robust to non-normal distributions, making it more appropriate for testing whether higher levels of perceived importance of safety or liveability, for example, tended to co-occur with larger proportions of specific street elements such as vegetation or cycle lanes (de Winter et al., 2016). These results were valuable for hypothesis-building and identified design to perception links that could later be formalised and assessed within the multivariate framework. The lack of several significant correlations in this stage, shown in

Appendix A, also suggested that image-derived variables by themselves weren't sufficient to fully explain perceptions, justifying the need for moderation in later analysis stages.

Finally, associations between categorical and ordinal variables with preference were explored using mosaic plots. These plots not only confirmed where relationships existed but also revealed which categories were driving them through standardised residuals (Hartigan & Kleiner, 1981). For example, Figure 22 shows that being an urban specialist was found to be correlated with the importance given to safety when choosing between two street scenarios.

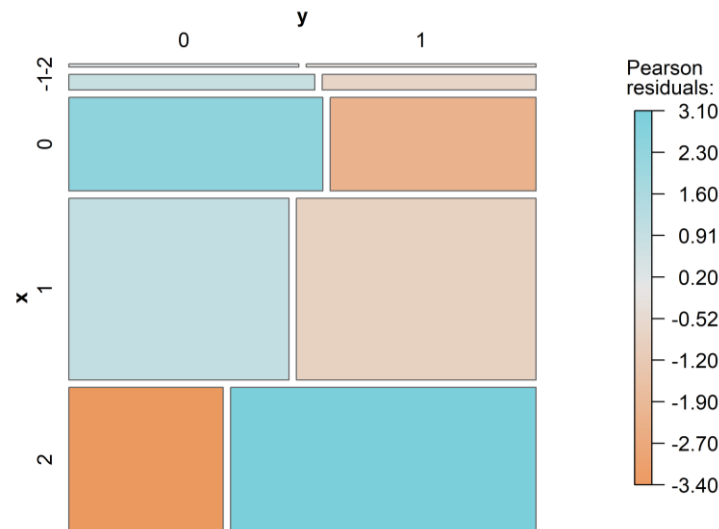


Figure 22 - Mosaic plot between x (safety) and y (urban specialist) variables.

The strong blue hue, in Figure 22, indicates overrepresentation (positive correlation) between being an urban specialist ($y = 1$) and the maximum importance being given to safety ($x = 2$), while the strong orange hue suggests underrepresentation (negative correlation) between not being an urban specialist ($y = 0$) and the maximum score being given to safety relevance ($x = 2$). This provided an early indication of how background, everyday practices and mobility habits might moderate preference over urban interventions, insights that were subsequently formalised in the structural models.

Overall, the bivariate analysis served more than a descriptive role. It was diagnostic and preparatory, flagging redundancy between predictors, identifying candidate mediators and moderators and clarifying how perception items grouped. These insights ensured that the subsequent models were not only statistically tractable, by reducing multicollinearity and dimensionality, but also empirically grounded in observed associations.

In short, while structural equation modelling (SEM) was the main analytical framework, bivariate analysis was indispensable in shaping its specification and ensuring its robustness as reported by Kline (2016).

3.7.3 Polychoric factor analysis on perception variables

The analysis of perception items required an additional step, besides the polychoric correlation since they would serve as indicators of the latent constructs in the SEM models. Thus, a measurement model was needed, which would define the paths between perception variables and these undetermined factors.

Before proceeding with factor extraction, data suitability was evaluated. Two common diagnostics were applied: the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity following (Smyth & Johnson, 2025) suggested procedure.

The KMO statistic evaluates whether the observed correlations are large relative to partial correlations. High KMO values indicate that the variance in the dataset can be explained by underlying latent factors rather than by noise. In simpler terms, a high KMO suggests that items group together and that factor analysis will meaningfully reduce the data. The result of this test was a 0.848 KMO value, which is considered "*Meritorious*" (Kaiser, 1974).

Bartlett's test, by contrast, checks whether the correlation matrix significantly differs from an identity matrix, in which case there would be no meaningful shared variance to explain. The significant result ($p\text{-value} < 0.05$) indicates that the observed variables are indeed correlated enough to justify factor analysis (Smyth & Johnson, 2025). These checks confirmed the data's suitability, enabling the following parallel analysis step.

Unlike the traditional Kaiser criterion, which keeps all factors with eigenvalues greater than one, parallel analysis generates random datasets of the same size and compares their eigenvalues to those of the real data (Lim & Jahng, 2019). Only those factors whose eigenvalues exceed the simulated ones are retained, ensuring that retained factors reflect genuine structure rather than sampling artefacts. The results were visualised in a scree plot, shown in Figure 23, where the line of actual data eigenvalues was compared to the simulated benchmark.

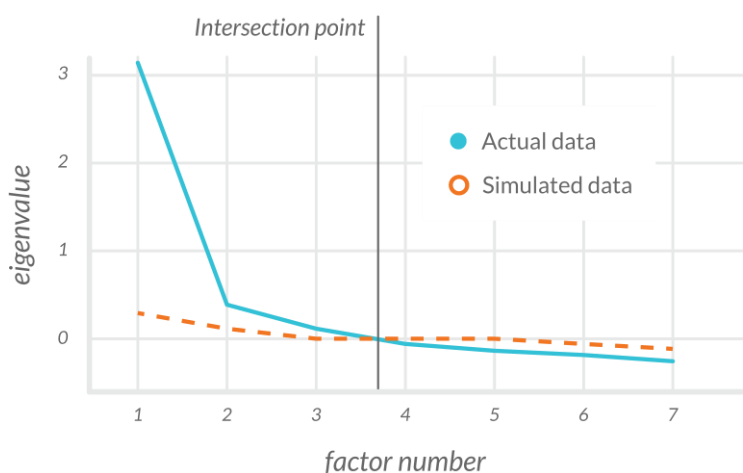


Figure 23 - Parallel analysis' scree plot.

With these checks complete, exploratory factor analysis could be computed. EFA is used when the latent structure of the data is not firmly established and the goal is to uncover patterns in how items cluster together without theoretical bias (Fabrigar et al., 1999). Principal axis factoring (PAF) was used with oblimin, an oblique rotation method that allows extracted factors to correlate. This choice reflects the theoretical expectation that psychological and perceptual constructs as safety, comfort and liveability, are not independent dimensions but interconnected (Costello & Osborne, 2005). A simple analogy would be that a neighbourhood that feels unsafe might also feel stressful or that a place perceived as functional may also seem habitable. Forcing independence through an orthogonal rotation, such as varimax, would therefore misrepresent these natural relationships (Costello & Osborne, 2005).

Complying with what the parallel analysis suggested, in Figure 23, two EFA solutions were tested,

one with two factors and one with three. This process is illustrated in Figure 24, for better visualization of the abstract methods. For each solution, several measurements were examined: factor loadings, which indicate how strongly each observed item is associated with a factor; communalities, which show how much of each item's variance is explained by the factors; eigenvalues, which provide information on the importance of each factor; and factor correlations, which reveal how closely the latent constructs are related.

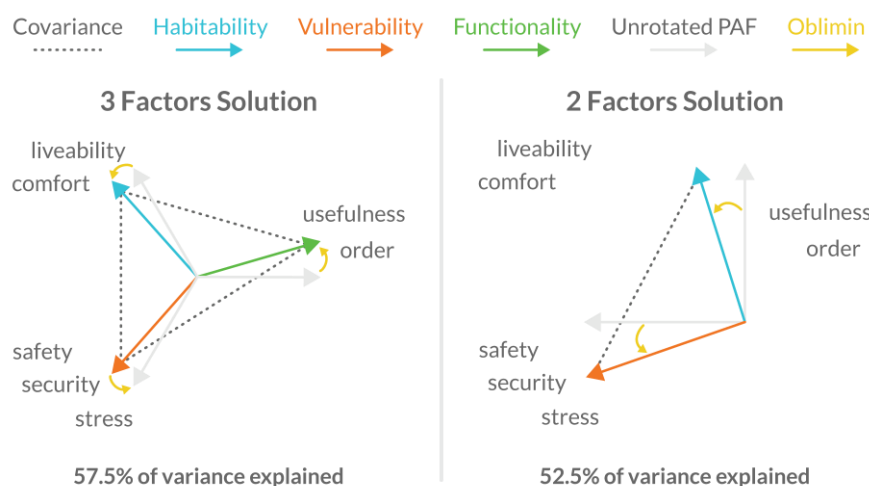


Figure 24 - Representation of the EFA procedure on perception variables.

The two-factor solution suggested a broad division of perceptions, but the three-factor solution was more interpretable. The latter separated perceptions into three coherent domains: functionality, vulnerability and habitability. Besides, as shown in Figure 24, while the two-factor solution explained around fifty-three per cent of the variance, the three-factor solution explained around fifty-eight per cent.

Having evaluated both preliminary structures, the confirmatory factor analysis (CFA) was conducted. Whereas EFA is inductive and exploratory, CFA allows testing of a specific hypothesised structure against the observed data (Fabrigar et al., 1999). Based on the EFA results, a three-factor model was chosen in which usefulness and order were indicators of “Importance of Functionality”, stress, safety and security defined “Importance of Vulnerability” and comfort and liveability represented “Importance of Habitability”. It is relevant to point out that despite these factors being mentioned by their simpler terms throughout this dissertation, for a better reading experience, (functionality, habitability and vulnerability), they in fact represent the importance given to those terms when choosing between the before and after images. The same happens with their indicators (usefulness, stress, comfort, liveability order, security and safety) which were rated on a 5-point Likert scale when asked “*How much did these terms influence your choice?*” (Lopes, 2025). This is an important distinction between later predicting perceptions felt and predicting the importance of said perceptions.

Correlations between the three latent factors were freely estimated, reflecting again that these constructs are theoretically related. The CFA was estimated using the WLSMV estimator, which is designed for ordinal categorical indicators and is more robust than maximum likelihood when dealing with non-continuous data (C. H. Li, 2016). The model produced parameter estimates, standardised factor loadings, factor scores for each respondent and global model fit indices.

Model fit was then assessed using these standard indices, each capturing a different aspect of how

well the three-factor structure reproduced the observed data. The chi-square test yielded a value of 22.02 with 11 degrees of freedom, corresponding to a p-value of 0.024. Although this result is formally significant, chi-square is well known to be overly sensitive to sample size and even small deviations from a perfect fit can produce significance. For this reason, greater emphasis is placed on alternative indices, as suggested by the literature (Hu & Bentler, 1999).

The comparative fit index (CFI) was 0.999 and the Tucker–Lewis’s index (TLI) was 0.998, both comfortably above the conventional 0.95 threshold and indicating that the hypothesised model explained the data perfectly compared to a null independence model (Hu & Bentler, 1999). The root mean square error of approximation (RMSEA) was 0.029, well below the 0.06 cut-off that signals close fit and the standardised root mean square residual (SRMR) was 0.021, again well within the acceptable range of 0.08 (Hu & Bentler, 1999). Taken together, these indices provide strong evidence that the three-factor specification fits the data extremely well.

This stage confirmed that the relevance of perceptions of the street environment can be structured into three distinct but correlated latent constructs, pictured in Figure 25.

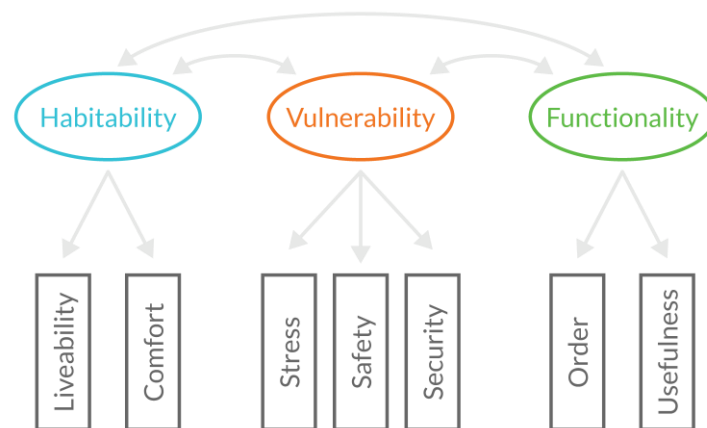


Figure 25 – Latent construct scheme based on the EFA and CFA processes.

Functionality captured whether a street was perceived as useful and orderly, Vulnerability reflected whether it evoked stress and feelings of insecurity and Habitability represented whether it was seen as comfortable and liveable. They are distinct enough to warrant separate treatment yet interconnected in ways that mirror the complexity of lived experience. The structure derived from this model therefore provides a solid, theory-driven and statistically reliable foundation for the subsequent predictive modelling of public preference on street interventions.

3.8 Structural equation modelling

The final analysis stage aimed to connect the objective information extracted from street images with the subjective evaluations provided by survey respondents. Having established reliable latent constructs through confirmatory factor analysis, the next step was to examine how these would mediate the paths between segmentation and colour predictors to the overall preference outcome. SEM was adopted as the central framework because it permits the simultaneous estimation of measurement and structural components, reduces measurement error and accommodates both direct and indirect effects, while still allowing for moderation (Marôco, 2014). This chapter first outlines the construction of the base models for

segmentation and colour predictors, before turning to the moderation models that analyse whether these relationships differ systematically across demographic and commuting groups.

However, the SEM analysis first required preparing the survey and image-derived variables. Once more, each variable type needed different treatment so all variables could be used and interpreted together.

Besides, during preliminary testing the variable for road space (road_spc_var) was found to exhibit strong multicollinearity with all segmentation predictors. This could be explained by the fact that the sum of the variation represented by each segmentation variable must always be zero. When road space increases another segmentation variable must decrease, since the before and after images have the same resolution (number of pixels). To prevent coefficient inflation and sign inversion, road space was excluded from the SEM analyses, becoming a reference variable for their interpretation.

Continuous predictors were standardised into z-scores. This scaling step ensured that variables measured on different numerical ranges contributed comparably to the estimation process, thereby avoiding distortions in coefficient size. Standardisation also facilitated the later interpretation of effect sizes, since coefficients could be read in terms of standard deviations (Atif, 2024).

Binary variables such as familiarity were converted into factors, making them suitable for SEM estimation in R, while multi-level demographic variables were flagged for dummy coding in later moderation models.

A specific adjustment was also necessary for the colour variables, since the raw measures of brightness and saturation, including both means and standard deviations, were found to be strongly correlated with one another in the bivariate analysis results. To reduce redundancy and simplify interpretation, a principal component analysis (PCA) was conducted with two retained factors, rotated by varimax to maximise loading separation as proposed by Kavengik (2025). The resulting dimensions were labelled vividness–saturation and contrast–brightness, which captured complementary aspects of colour perception. Factor scores for these two components were saved to the dataset and used as predictors in the SEM models, while the original four brightness and saturation indicators were dropped to prevent collinearity. This dimensionality reduction allowed colour information to enter the models in a parsimonious yet theoretically meaningful way. To allow for a better understanding of the PCA factors, a schematic plot is presented in Figure 26.

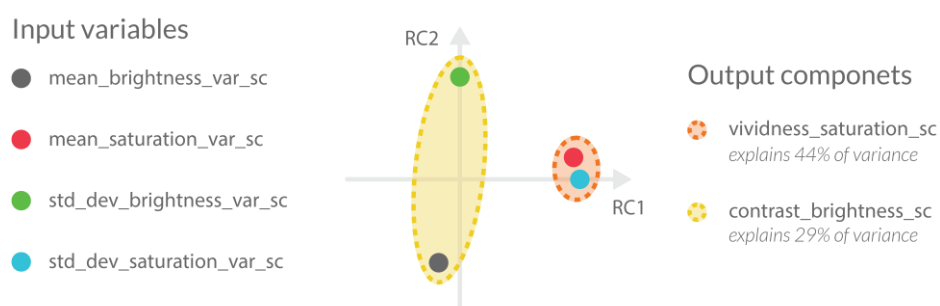


Figure 26 - Representation of the PCA on brightness and saturation variables.

3.8.1 Base models

Due to a substantial number of possible combinations between predictors and moderators (340

different interactions requiring a sample size above 667 100 observations, according with Equation 4), two base models were first defined, one for segmentation-derived predictors and another for colour-derived ones, before moving into demographic moderation. The two kinds of predictors were separated into two base models due to their different theoretical natures. While segmentation represents the physical objects in a street scene, colour represents the visual aspects of those objects. Despite this theoretical distance between the two concepts, it was possible to detect several strong correlations between segmentation and colour variables, sometimes from confounding items, which would've caused collinearity if only one base model were used. Some of these identified correlations are easily understood by noticing that some street objects, such as bus lanes, commonly have similar visual representation across cities and streets as a means of facilitating identification. Thus, it is possible to analyse, in Appendix B, that red colour presence variation was found to have a strong positive correlation with bus lane space variation (Pearson's $r = 0.80$, $p\text{-value} < 0.05$).

To define these base models, two components were needed: the measurement model, which defines how observed variables (perception indicators) represent underlying latent constructs; and the structural model, which specifies the relationships between those latent constructs, predictors and the outcome item, to evaluate hypotheses about cause and effect.

The measurement model was based directly on the constructs validated in the confirmatory factor analysis, depicted in Figure 25. Functionality was defined as the latent factor underlying perceptions of order and usefulness, Vulnerability was represented by safety, security and stress and Habitability combined comfort and liveability. These three latent constructs were permitted to correlate with one another, acknowledging again that perceptions of the same environment are not independent but intertwined.

On top of this measurement model, the structural models were defined and illustrated (example in Figure 27). In both cases, all segmentation-derived or colour-derived variables were specified as predictors of Functionality, Vulnerability and Habitability, as well as of the binary outcome variable Preference. The only exception being the road variable, which was omitted from these predictor sets due to the previously identified multicollinearity effects. The three latent constructs were included as mediators predicting preference, ensuring that the models could capture both direct and indirect pathways from objective features to subjective outcomes.

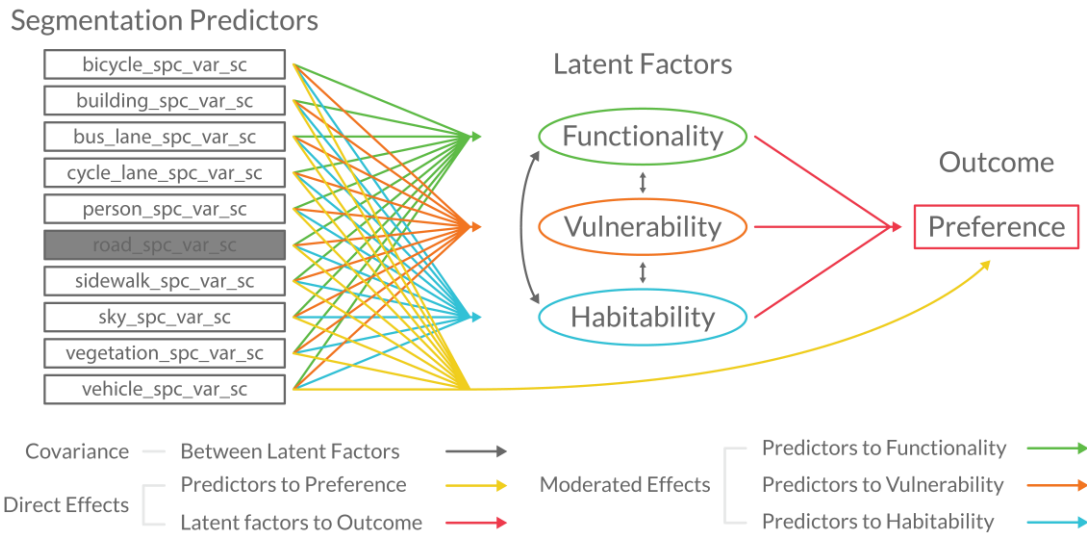


Figure 27 - Representation of the segmentation-derived base SEM model.

Model estimation was conducted using the WLSMV estimator, which is recommended when dependent variables are ordinal, as is the preference variable (C. H. Li, 2016). For each model, standardised parameter estimates, fit measures, modification indices and R^2 values were extracted and exported both in full form for validation and in reduced form for integration into the predictive tool.

Together, these base models provided the essential paths between image-derived predictors and perception outcomes. They established the degree to which segmentation and colour information could explain variation in the latent constructs and preference ratings, thereby laying the empirical foundation for subsequent moderation analyses that explored how these relationships differed across population subgroups.

3.8.2 Moderation models

The moderation stage of the analysis extended the base SEM models by examining whether the relationships between objective predictors, latent constructs and stated preference varied across different demographic groups.

However, before such interactions could be evaluated, the demographic variables required careful preprocessing to ensure comparability and interpretability. Variables such as age, commute time and the number of commute modes were first grouped into categorical bins. Age was recoded into three brackets labelled “Young,” “Middle,” and “Older,” capturing early-career (< 29), mid-career (29 - 49) and later stages (> 49) (Arnett, 2000). Commute time was condensed into the categories “No commute,” “Short,” and “Long,” while the number of commute modes was represented as “None,” “Single-mode,” “Two-mode,” and “Multi-mode.” These transformations provided meaningful categorical distinctions that aligned with theoretical expectations about lifestyle differences, while also making the data more suitable for interaction modelling.

Multi-level demographic variables were then converted into dummy variables so that their effects could be modelled without imposing unwarranted ordering and the first category of each variable was dropped to serve as the reference group.

Gender and occupation, however, required specific handling. Gender was restricted to male and

female responses, with “other” being excluded due to insufficient representation noticed during the univariate analysis and recoded into a binary variable with females coded as one. Occupation was simplified into employed versus not employed, combining students, retired individuals and the unemployed into a single reference group, again due to insufficient representation of these subgroups.

Once the demographic variables had been pre-processed, moderation effects were computed through residual-centred interaction terms, a process simplified in Figure 28.

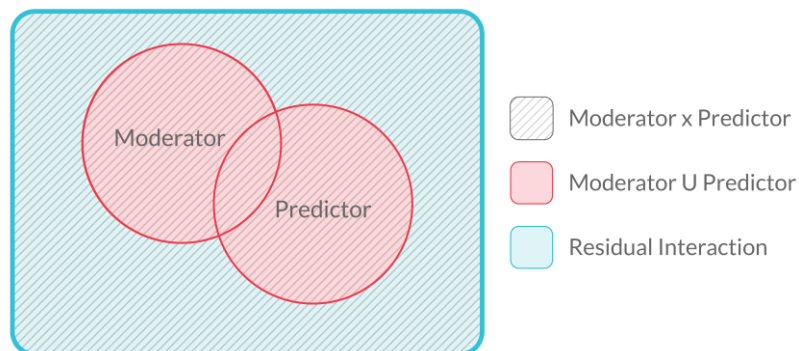


Figure 28 - Representation of the residual-centred interaction term process.

For each predictor-moderator pair, a raw interaction was constructed by multiplying the two variables together. This raw term was then regressed on the main effects and the residuals from that regression were extracted as the final interaction term. This residual-centring procedure preserves the interpretability of the main effects, which remain unconfounded by the interaction and avoids the severe multicollinearity that arises when interaction terms are included directly (Geldhof et al., 2013). To ensure reproducibility in the predictive tool, the coefficients from each centring regression were also stored and exported alongside the moderation results.

Due to the high number of interaction term parameters that needed to be estimated (around 1 462), the SEM moderation analysis proceeded in three structured loops (one iteration parameter at a time), each representing a different pathway through which demographic characteristics could shape perceptions and preference. A simplification of the moderation SEM setups (same structure for segmentation-derived variables as for colour-derived variables) is presented in Figure 29.

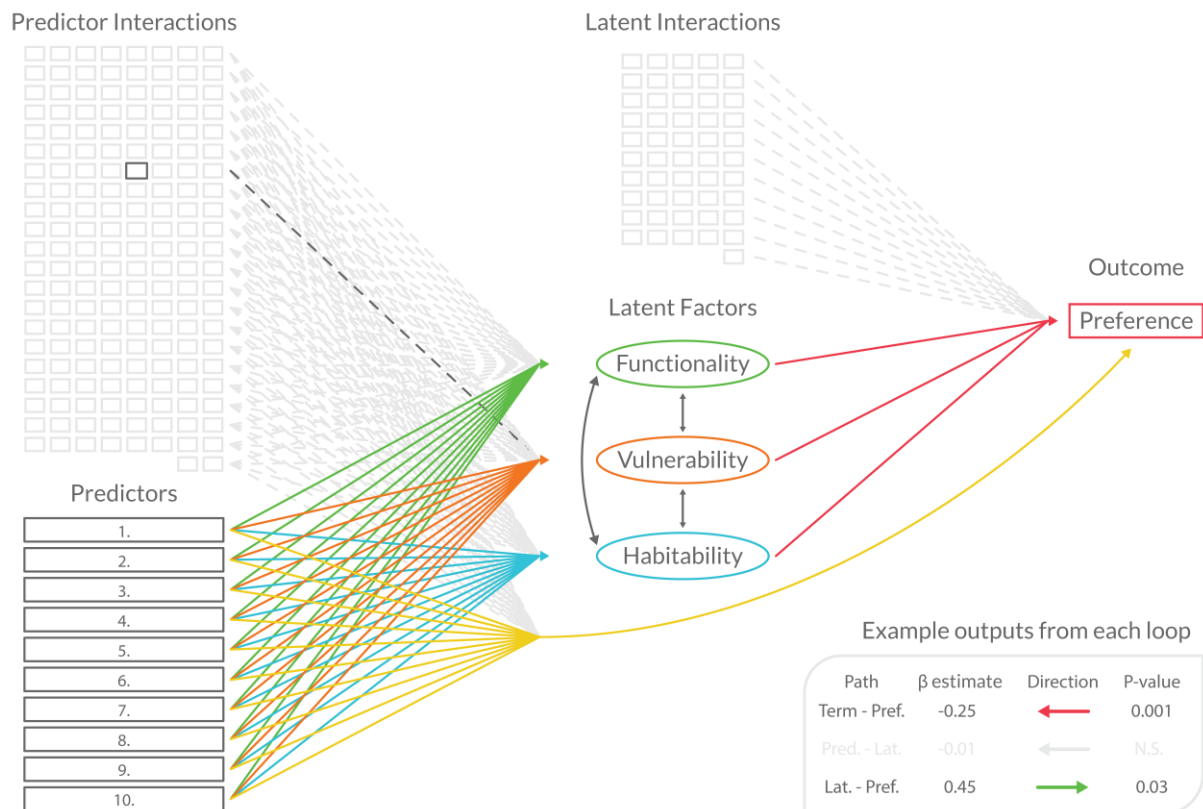


Figure 29 - Simplification of the moderation SEM for colour or segmentation predictors.

The first loop assessed interactions between observed predictors and demographics in predicting the latent constructs of Functionality, Vulnerability and Habitability. This captured, for example, whether the effect of vegetation on perceived importance of habitability varied by age group or gender.

The second loop evaluated interactions between predictors and demographics in directly predicting preference over the mediating role of the latent constructs. This allowed the models to detect cases where a street feature might be particularly appealing or unappealing to certain demographic groups, regardless of how it influenced perceptions of Functionality, Vulnerability or Habitability.

The third and final loop assessed interactions between these latent constructs themselves and demographics in predicting preference. In this specification, the focus was on whether the weight given to a particular perception differed across subgroups. For instance, whether respondents with children placed more importance on Vulnerability regardless of what street environments they preferred.

For each interaction parameter assessed, a separate SEM system was estimated using the WLSMV estimator. This term-by-term estimation strategy was deliberate since including all interactions simultaneously would have increased model complexity, going from 106 total estimate paths to 1 568 and risked overfitting, particularly given the moderate sample size of 1 203 observations. According to [Westland \(2010\)](#), around 667 100 observations would be needed (Equation 4), if the standard SEM approach was chosen for this dissertation, which includes all predictors, latent constructs, respective indicators, outcome variables and interaction terms in a single model.

$$N > 50 r^2 - 450 r + 1100$$

Equation 4 - Minimum sample size for SEM. ([Christopher Westland, 2010](#))

By estimating models one interaction at a time, the procedure ensured computational feasibility and made it possible to attribute significant moderation effects to specific predictor x moderator combinations. Although this approach does not capture higher-order dependencies among moderators, like the traditional singular SEM workflows described by [Marôco \(2014\)](#), it provides a transparent and interpretable way of mapping demographic contingencies in the data.

Alternative solutions would have needed a dimensionality reduction step of demographic variables or the binning of respondents based on their information. Both options would have created black box scenarios where the interpretability of what measures impact preference is not possible. Transparency and interpretability were therefore chosen over better model fit estimates.

The output from each model, simplified in Figure 29, included standardised estimates, standard errors, z-values and p-values for the interaction terms, alongside model fit statistics such as the chi-square test, degrees of freedom and associated p-value. With these in mind, interaction effects were then flagged as significant ($p\text{-value} < 0.05$) or not significant ($p\text{-value} \geq 0.05$).

Additionally, other outputs were produced to support the reproducibility of preprocessing steps in the predictive tool and interpretation of results. PCA results for the colour variables were summarised in a minimal workbook, including loadings, eigenvalues, variance explained and descriptive statistics of the original variables. Mapping rules for gender and occupation were saved in separate files. Means and standard deviations of all continuous variables were recorded for later use in variable scaling. Finally, the residual-centring coefficients for all interaction terms were also exported, enabling new datasets to be adapted to the training data's characteristics.

The whole moderation framework allowed for a systematic exploration of how respondent's characteristics and habits condition the pathways from street features and colour environments to the importance of perceptions and stated preference. By structuring the analysis around three distinct loops and maintaining a strict separation between segmentation and colour predictors, the models revealed not only whether predictors mattered overall, but also whether and how their effects differed across community descriptors. The results from this final analysis form the structure needed to interpret and predict socially differentiated responses to street-level interventions, thereby strengthening the later described predictive tool's ability to capture the diversity of public reactions to urban change.

3.9 Predictive tool

Lastly, after the full univariate, bivariate and SEM analysis, a predictive tool was designed to estimate the probability of a group of respondents preferring a proposed street design based on the coefficients derived from the structural equation models. To ensure out-of-sample evaluation, the predictions were always generated on a holdout dataset, which had not been used in the estimation of either the base or the moderation models.

All preprocessing steps were reproduced on the holdout data using only parameters obtained during training, guaranteeing consistency between the datasets and avoiding any form of information leakage. For this, continuous predictors were standardised using the means and standard deviations stored. For the colour-based model, the two principal components capturing vividness/saturation and contrast/brightness were reconstructed by applying the saved loadings to the holdout data, thereby

ensuring that the latent colour factors matched those used in training. Demographic variables were re-coded in the same way as during model estimation: gender was transformed into a binary with females coded as one and responses marked as “other” excluded; occupation was recoded so that employed respondents were coded as one and students, retired individuals or those unemployed were grouped as zero. Age, commute time and commuting mode were binned in line with the moderation rules applied previously. Lastly, categorical demographic variables were dummy coded, making sure to align their respective names and dropped dummies with the one used in training.

To obtain higher fidelity results, prediction relied only on the subset of parameters that had been automatically filtered as statistically significant at the p -value < 0.05 threshold, unless a moderation path was significant without the related predictor or latent factor itself being significant in the base models. In this case, both the base and moderation path estimates were used to maintain theoretical coherence. This automatic filtering approach allows for a more adaptable workflow, permitting different communities with varied significant factors to use the exact same process. It also protects new models' structures from bias towards the data used in this dissertation by allowing the structure itself to adapt to new trained data, based on statistical significance. The combination of the term-by-term moderation system with the automatic significance flagging is what made reproducibility across any community feasible.

To further calibrate the prediction tool an extensive cross-validation framework based on intervention-specific holdout groups was computed. Each holdout group corresponded to one intervention, defined by its survey block codes. When an intervention appeared in more than one block (e.g., 301 and 405 are the same intervention in blocks 3 and 4), both code sets were held out together, ensuring that all associated data for that intervention was excluded simultaneously. This prevented information leakage between training and testing subsets. To preserve sufficient sample size for model estimation, a small number of interventions essential for structural stability were excluded from being holdouts and were always included in the training data (101, 103, 104, 202, 402 and 407).

The model was trained iteratively on 24/25 groups at a time, depending on whether the current holdout contained one or two intervention codes, repeating this process across 17 holdout groups. After each training iteration, the predictive tool estimated preference probabilities for the corresponding holdout data, which were then compared to the observed preference proportions for that intervention. The difference between the mean predicted probability and mean observed preference was computed as the Error of Mean, representing population-level accuracy, while the absolute errors averaged across all 17 holdouts yielded the Mean of Errors, reflecting the average deviation per intervention. Together, these indicators captured both global calibration and group-level precision.

To determine optimal calibration, shrink factors were applied systematically to the model's coefficients. These shrink factors were constants that scaled the influence of each estimate on the t_1 threshold, which defines the baseline preference probability. For the initial exploration, shrink combinations ranged from 0 to 1 in 0.1 increments for both base and moderation coefficients (11 by 11 grid). Each combination was evaluated across all 17 holdouts, producing 2 057 predictions in total. The Error of Mean and the Mean of Errors were then represented in a colour coded table per predictor type, to facilitate the extraction of the best shrink combination for both the colour and segmentation systems. The colour-related results were depicted in Table 3.

Table 3 - Shrink factor tuning results for the colour-related workflow.

Colour		Base Estimates																					
		0		0.1		0.2		0.3		0.4		0.5		0.6		0.7		0.8		0.9		1	
Moderation Estimates	0	8.1%	19.4%	7.7%	18.9%	7.2%	18.9%	6.5%	18.9%	5.7%	18.9%	4.7%	19.1%	3.6%	19.4%	2.3%	20.1%	0.9%	20.9%	0.6%	21.6%	2.2%	22.3%
	0.1	8.0%	19.4%	7.7%	19.0%	7.1%	18.8%	6.5%	18.8%	5.6%	18.8%	4.7%	19.0%	3.5%	19.3%	2.3%	20.0%	0.9%	20.7%	0.6%	21.5%	2.1%	22.2%
	0.2	7.9%	19.5%	7.5%	19.1%	6.9%	18.9%	6.3%	18.8%	5.4%	18.7%	4.5%	18.9%	3.4%	19.2%	2.2%	19.8%	0.8%	20.6%	0.6%	21.3%	2.2%	22.0%
	0.3	7.6%	19.7%	7.2%	19.3%	6.7%	19.1%	6.0%	19.0%	5.2%	18.9%	4.2%	18.8%	3.1%	19.0%	1.9%	19.6%	0.6%	20.4%	0.8%	21.1%	2.3%	21.9%
	0.4	7.3%	19.9%	6.9%	19.6%	6.3%	19.4%	5.6%	19.3%	4.8%	19.2%	3.8%	19.2%	2.8%	19.3%	1.6%	19.7%	0.3%	20.2%	1.1%	20.9%	2.5%	21.6%
	0.5	6.9%	20.1%	6.5%	19.8%	5.9%	19.6%	5.2%	19.6%	4.4%	19.5%	3.4%	19.6%	2.4%	19.7%	1.2%	20.0%	0.1%	20.6%	1.4%	21.1%	2.8%	21.6%
	0.6	6.6%	20.2%	6.1%	20.0%	5.5%	19.9%	4.8%	19.8%	3.9%	19.8%	3.0%	19.9%	1.9%	20.1%	0.8%	20.4%	0.5%	20.9%	1.8%	21.4%	3.2%	21.9%
	0.7	6.2%	20.4%	5.7%	20.2%	5.1%	20.1%	4.3%	20.1%	3.5%	20.1%	2.5%	20.2%	1.5%	20.5%	0.4%	20.7%	0.9%	21.2%	2.2%	21.8%	3.5%	22.3%
	0.8	5.8%	20.5%	5.3%	20.4%	4.7%	20.3%	3.9%	20.3%	3.1%	20.4%	2.1%	20.5%	1.1%	20.8%	0.1%	21.1%	1.3%	21.6%	2.6%	22.1%	3.9%	22.6%
	0.9	5.5%	20.6%	4.9%	20.6%	4.3%	20.5%	3.5%	20.6%	2.7%	20.6%	1.7%	20.8%	0.7%	21.2%	0.5%	21.4%	1.7%	21.9%	2.9%	22.4%	4.3%	22.9%
	1	5.1%	20.7%	4.6%	20.7%	3.9%	20.7%	3.1%	20.7%	2.3%	20.8%	1.3%	21.1%	0.3%	21.5%	0.8%	21.7%	2.0%	22.2%	3.3%	22.7%	4.6%	23.2%

The blue and purple range in Table 3 represented the Error of Means of all shrink combinations, light blue being the lowest and purple being the highest result. The green to red range evaluated the Mean of Errors of all shrink combinations from lowest to highest, respectively. Table 3 also includes a black rectangle which was used to mark the best results, therefore used in later predictions.

For segmentation, the optimal combination appeared at the lower-right corner of this initial grid, suggesting that the model could benefit from higher shrink values than those initially assessed. Consequently, the 11 by 11 grid for segmentation was expanded to cover shrink factor combinations between 0 and 3 in 0.1 intervals (31 by 31 grid), yielding 961 combinations each evaluated across all 17 holdouts (16 337 predictions in total). In contrast, the colour-based system stabilised around mid-range values on the first exploration, reflecting different coefficient distributions and sensitivity to multicollinearity between the two types of predictors.

Based on these evaluations, segmentation-derived coefficients were shrunk using 1.2 for base estimates and 0.9 for moderation, while colour-derived coefficients used 0.8 for base and 0.4 for moderation. These shrink settings were fixed and applied to all subsequent real predictions to ensure consistent calibration across both systems. These values produced the lowest combined Error of Mean and Mean of Errors (segmentation: 0.01% / 9.70%; colour: 0.30% / 20.24%). As expected, the Error of Mean remained consistently lower than the Mean of Errors, confirming that SEM predictions are highly accurate at the population level, while individual-level predictions remain more variable.

The different optimal shrink factor combinations for colour and segmentation are consistent with the characteristics of each model: colour predictors tend to have more dispersed coefficients and higher variance which required more control over estimates, whereas segmentation predictors are tightly linked, requiring less aggressive shrinkage to prevent coefficient inflation and multicollinearity effects.

To create the interactions terms for each holdout prediction, residual-centring was also used to reproduce the same scaling procedure as in the training phase. Each interaction term was recomputed

in the holdout data by first multiplying the item and moderator and then subtracting the fitted values obtained from the stored regression coefficients of the interaction residualisation procedure. This ensured that the interactions were measured on the same residualised scale used when fitting the SEMs, thereby maintaining compatibility with the path estimates.

Once the predictors and interaction terms were reconstructed, latent factor scores were computed deterministically as linear combinations of their relevant predictors and interactions weighted by the corresponding coefficients. For each latent construct, this meant summing the products of the retained predictor values with their associated coefficients and adding the contributions of any significant moderated effects. The resulting factor scores represented the model-implied importance of each perceptual dimension for the new observations. This process is also described in a simpler visual way in Figure 30.

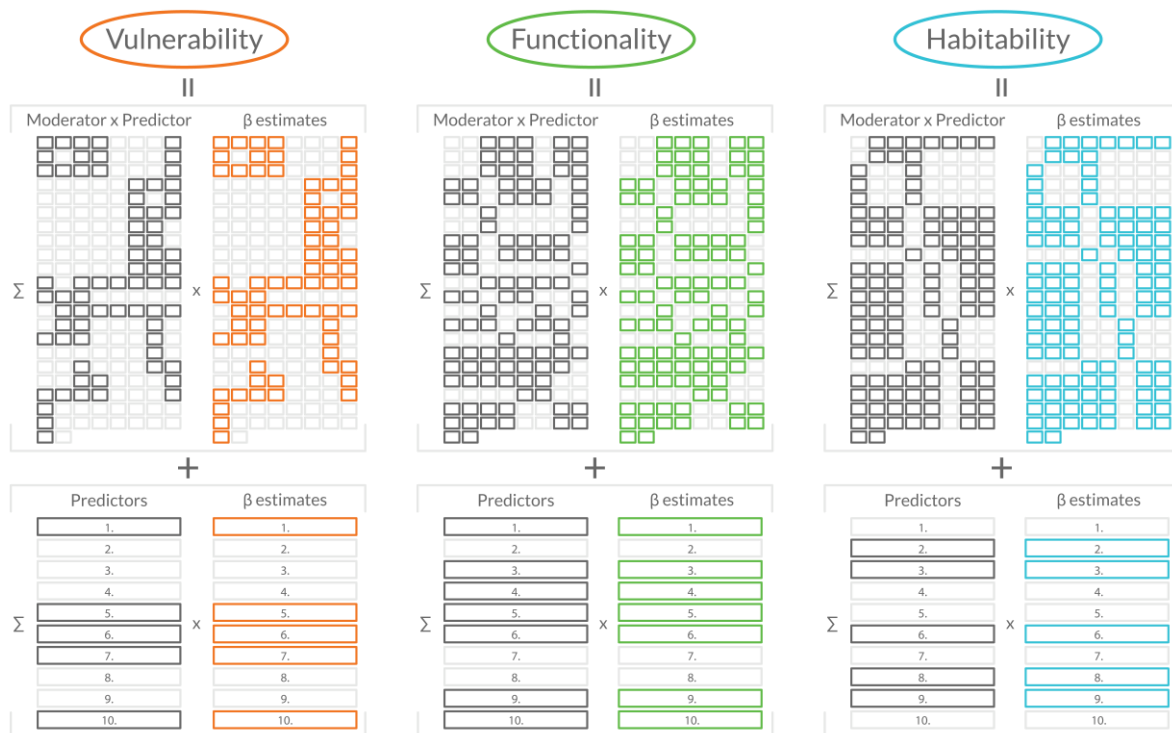


Figure 30 - Representation of the latent construct prediction process.

It's important to note that the greyed-out cells, in Figure 30, represent the non-significant predictor' and interaction term' estimates made by the training models. Therefore, these were not included in the latent construct predictions. Additionally, terms could be significant for certain latent constructs without being significant for all of them. Adding this challenge to the fact that there were two separate predictions per latent construct (one for colour-derived models and another for segmentation-derived models) heightened the complexity of computation but did not affect the simplicity of results. From this latent construct prediction step, the tool outputted two scores per latent construct per respondent.

The Preference outcome was then predicted by combining the latent factor scores and any significant direct predictors or interaction terms into a similar linear predictor. This linear predictor was subsequently converted into a probability using the base term estimated in the SEM.

In other words, the probability that Preference would equal one was given by applying the normal cumulative distribution function to the linear predictor, shifted by the estimated threshold parameter t_1 .

Consequently, the predictive tool delivered two calibrated probability scores for each respondent in the holdout dataset (one from the colour-derived workflow and second one from the segmentation-derived framework). By applying the procedure separately to these segmentation and colour-based models, predictions could be compared across different feature spaces. In both cases, all steps from scaling, factor reconstruction, dummy coding, interaction computation, coefficient selection and probability estimation were deterministic and anchored in the original training artefacts, guaranteeing full reproducibility.

The predicted probabilities could then be summarised, visualised in histograms and saved alongside parameter effect plots that illustrate the contribution of significant predictors and interactions. Together, these steps confirmed that the validated SEM structures could be used not only to understand the drivers of perceived Functionality, Habitability and Vulnerability, but also to provide an empirically grounded predictive framework for estimating how respondents evaluate specific street-level changes.

3.10 Demonstration of applied value

This section demonstrates the application of the predictive framework developed in this thesis to a real-world scenario, illustrating how the method could be operationalised to evaluate five alternative interventions for a single street segment, Avenida Rio de Janeiro, in Lisbon. The aim was not to assess which design was “best,” but to show how the tool integrated the relationships identified through the Structural Equation Models (SEM) into a reproducible workflow for perception prediction.

The predictive tool translated the statistical pathways linking physical street features, perceptual constructs and overall preference into a computational structure where each SEM coefficient defined the expected influence of a given change on perceived importance of Functionality, Habitability, Vulnerability and overall Preference. When applied to new visual data, this framework generated probabilistic estimates representing how different respondent groups might perceive a proposed redesign.

To create these hypothetical intervention scenarios, presented in Figure 31, a combination of AI and manual image editing was used. DALL·E, an image generation model by OpenAI, was chosen as the main generation/modification model due to its open copyright policy. Using artificial intelligence (AI) to imagine the street's different urban interventions minimized bias towards known estimates, with the only condition being that each intervention should focus on different mobility modes, mirroring the survey approach from (Lopes, 2025). An additional “MixUse” intervention was also generated to evaluate results on a more diverse scenario. The manual editing step, in Adobe Illustrator, ensured that images were coherent (e.g. removing people from the road and adapting wide cycle lanes into a more standard size).

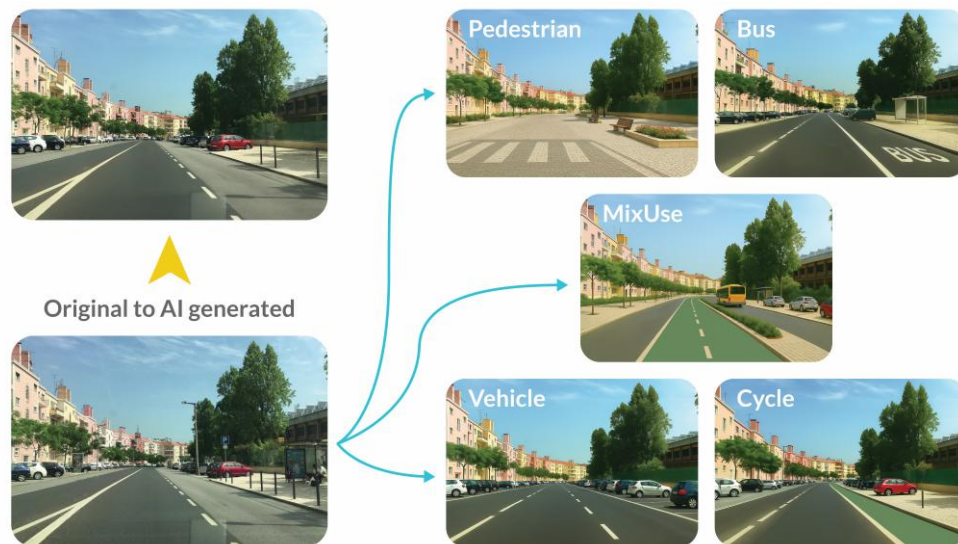


Figure 31 - Alternative interventions to Avenida Rio de Janeiro, Lisbon.

As shown in Figure 31, the original image from KartaView was also passed through DALL·E to decrease artefacts from comparing generated images with real world photographs. This created a blank slate of the street which could then be compared to the different hypothetical interventions.

The Pedestrian scenario removed traffic and introduced wider sidewalks, street furniture and vegetation, representing a walkability-focused redesign. This transformation increased the proportion of warm neutral tones and greenery while slightly reducing chromatic diversity due to the homogeneity of paving materials. The Bus intervention prioritised public transport by designating a dedicated bus lane and adding clear surface markings and signage, generating high-contrast zones where bright lettering and asphalt colour differences accentuated functional hierarchy and legibility. The Cycle configuration introduced green cycle lanes and clear spatial separation from cars, aligning with active-mobility objectives and producing a strong hue signature that was particularly relevant for colour-based perception analysis. The Vehicle scenario expanded traffic capacity and parking availability, representing a car-oriented layout dominated by darker grey values and lower overall saturation. Finally, the MixUse intervention combined elements of public transport and cycling infrastructure, achieving a balanced allocation between modes and a visually diverse but functionally integrated design that contained intermediate saturation and contrast levels.

Across all AI-generated scenarios, global contrast and colour saturation increased relative to the original photograph, reflecting the typical behaviour of image-generation models, which tend to enhance local contrast and introduce subtle noise patterns. These differences were not corrected, as they provided a consistent basis for analysing perceptual sensitivity to chromatic and luminance variation. Accordingly, each intervention simultaneously represented a functional configuration (captured by the segmentation model) and a visual signature (captured by the colour system), enabling a dual analysis of how spatial allocation and visual expressiveness jointly shape perceived Functionality, Habitability, Vulnerability and overall preference.

By analysing diverse designs within the same physical context, the approach isolated the perceptual effect of street features from contextual variability, allowing a clear view of how incremental physical modifications shift public perception. This approach more closely represents the real-world use of public

participation tools instead of evaluating different scenarios across different locations.

Before running these predictions, both the colour and segmentation base and moderation models were retrained on the full survey dataset, incorporating all available observations to ensure maximum stability and representativeness of the final coefficients. Each retrained model's estimates served as the foundation for predicting how respondents would evaluate alternative redesigns of the same street segment.

Afterwards, the prediction tool was computed on each intervention using the optimal shrink factors identified during the calibration stage in chapter 3.9 Predictive tool: 1.2 for base and 0.9 for moderation coefficients in the segmentation model; 0.8 for base and 0.4 for moderation coefficients in the colour model. These values were found to minimise prediction error and control coefficient inflation during cross-validation, ensuring that the final predictions maintained a balance between accuracy and interpretability.

This framework also replicated every preprocessing step described in the training and predictive tool methodology to ensure deterministic and reproducible outputs. Continuous predictors were standardised using the stored means and standard deviations from training. Demographic variables were recoded following the same categorical and binning logic used during model estimation, ensuring identical class definitions. Interaction terms were reconstructed using residual-centring based on the saved regression coefficients from the original interaction residualisation step, maintaining full consistency with the SEM measurement scale.

Following the tool's workflow, the latent constructs (Functionality, Habitability and Vulnerability) were calculated deterministically as weighted combinations of their significant predictors and corresponding moderated effects, enabling the same procedure to be computed for the overall preference outcome. For each step, only terms that had achieved statistical significance ($p\text{-value} < 0.05$) were included, while non-significant paths were omitted to preserve parsimony and avoid introducing noise. This process resulted in a coherent latent profile and preference score for each respondent under each proposed intervention.

Intervention-specific histograms of predicted preferences across users were then computed which helped in the quality control stage. Additionally, the tool exported a summary of the mean and standard deviation of preference probabilities and a full description file of all its calculation steps per user per intervention. The latter file enabled the management of thousands of data points per intervention into dynamic tables which allow users to compare and explore results across different combinations of demographic and habit groups.

4 Results and discussion

This chapter presents and discusses the outcomes of the framework developed in this dissertation, focusing on its capacity to generate, process and interpret visual and perceptual data through the proposed predictive tool. Rather than producing traditional standalone results, the aim is to demonstrate the tool's functionality, coherence and potential for application in evidence-based urban design.

Accordingly, the following sections validate each component of the workflow from image analysis to survey integration and predictive modelling, showing how they operate together to translate visual transformations of street environments into measurable preference. The first section assesses the behaviour and interpretative consistency of colour and segmentation metrics, confirming their suitability for large-scale visual data extraction. The second characterises the survey results that anchored perceptual inference and supported demographic moderation. The third applies the complete tool to a selected case study, illustrating its predictive capacity and interpretative depth. Collectively, these sections demonstrate how automated image analysis and perception modelling can be operationalised into a practical decision-support system, thereby addressing the research questions and core objective of this dissertation: the development and validation of a tool capable of linking visual street transformations with perceived urban quality.

4.1 Image analysis

The first stage of the results concerns the performance and applicability of the image-derived data gathering methods used, both in the colour and segmentation sections.

Although there is no procedure in the literature to quantify the quality of brightness, saturation, colour change or hue presence methods, every step described in chapter 3 Methodology followed standard best practices, applied across several studies ([Gorzaldini, 2016](#); [Habekost, 2013](#); [Jaglarz, 2023](#); [Luo et al., 2001](#)).

Additionally, the derived colour variables displayed heterogeneous distributions, with some concentrated near zero and others showing wider dispersion. The perceptual colour difference metric (`prctdeltaE_above_30`) had a mean of +23% and a median of +28%, indicating that in most cases a non-negligible proportion of pixels underwent changes above the perceptual threshold of $\Delta E_{2000} = 30$. The distribution was balanced (skewness = -0.05), suggesting that extreme cases of extremely high or exceptionally low perceptual change were uncommon.

In contrast, the hue-specific variance measures (e.g., `red_var`, `orange_var`, `yellow_var`) were characterised by highly skewed distributions, with most values close to zero and only a limited number of observations producing large positive deviations. This pattern reflects the fact that strong shifts in specific hues occurred rarely and were typically localised. For instance, `yellow_var` exhibited a particularly pronounced positive skew (+2.51), consistent with occasional but marked increases in yellow tones in certain projects such as the Barcelona superblocks, depicted in Figure 32.



Figure 32 - Barcelona superblock intervention example.

Brightness and saturation-based metrics showed broader variation across the dataset. The mean brightness variance was +8% but with values ranging from -23% to +45%, while mean saturation variance reached +28% on average but spanned an extensive range (-64% to +174%). In this context, percentage values for brightness and saturation refer to relative changes with respect to the before image, rather than absolute saturation on the HSV scale. These variables presented higher standard deviations compared to hue-specific measures, indicating that brightness and especially saturation were more consistently affected across projects.

Finally, distributional heterogeneity represented by `color_entropy_var` remained close to zero, with limited dispersion. This suggests that while the dominant hues and intensities could shift in individual cases, the overall diversity of colours in the images remained stable.

Having described the behaviour of colour-based metrics, the next section considers segmentation outputs. The main outcomes from the final training run, done during the semantic segmentation modelling stage, can be expressed in three key numbers. First, the validation loss reached a value of 0.1717 at the best checkpoint. This number reflects the average error between the model's predictions and the true labels. While lower values indicate a better fit, comparing them with other studies is not trivial, since it highly depends on the number of classes used (the higher the number the higher the chance of the model not learning one class well enough and lowering the average), which loss function was used and if it included weighted classes or not and the baseline performance of different datasets, pretrained weights and model setups.

However, to better understand the 0.1717 validation loss reached, a percentage of model “confidence” can be computed by inverting the cross-entropy function used (Equation 5):

$$CE(p, y) = -\log(p_y)$$

Equation 5 - Cross-entropy loss function.

, where p_y is the model's predicted probability for the correct class. If the model is “confident” ($p_y = 1$), then loss will be zero. The validation loss is the average of this value over all pixels in all validation images. By inverting Equation 5, it can be concluded that the model was $e^{-0.1717} = 82.4\%$ confident on average.

The second outcome from model training was the mean IoU across all classes in the validation set (0.4877). This means that when considering all nineteen classes together, the predicted segmentation regions overlapped with the ground truth regions by approximately 49% on average.

Lastly, the mean relevant IoU metric targeting specific categories of road, sidewalk, vegetation and

sky reached 0.5977. In practical terms, this indicates that the model's predictions for these four most frequent categories overlapped with the human annotations by 60% on average.

For comparison, efficient or real-time segmentation studies consider IoU metrics between fifty and seventy per cent on Cityscapes as a competitive threshold (Paul et al., 2020), while top of the line large models reach up to eighty-five per cent IoU (X. Liu et al., 2022).

After training was completed, the inference stage produced per-class space variation metrics (spc_var), representing the relative percentage difference in pixel share of each class between the before and after images. Their descriptive statistics reveal that most categories had distributions centred close to zero, which indicates that in most projects only small adjustments were predicted in terms of spatial allocation.

Road space, however, displayed a clear decreasing tendency, with a mean of -14% and a median of -11%, reflecting the systematic removal of car lanes in many projects, with the largest contraction reaching -59% in relative terms. Sidewalks showed the opposite trend, with positive means and medians (+10% and +3%, respectively) and a positively skewed distribution, suggesting that while many projects involved modest increases, a few introduced large reallocations of up to +34%. Buildings and vegetation were mostly centred around zero (means of +2% and 0%, respectively), but their long-tailed distributions suggest that changes in these categories often arise from occlusion effects, such as new trees covering façades, rather than from structural alterations to the streetscape. Sky also remained unchanged with a mean of -0.5%, although its distribution was negatively skewed, with reductions occurring in projects that introduced vegetation or elevated infrastructure.

Other categories showed more variable behaviour relative to their baseline prevalence. The number of visible people' pixels increased on average +1.2% but with a highly skewed distribution of 2.53, which indicates that although most scenes exhibited negligible differences, some showed substantial rises in pedestrian presence. A similar pattern was observed for bus lanes, which displayed rare but extreme increases of up to +59% (skew = 2.31). Vehicles, by contrast, tended to decrease (mean = -4.3%), reflecting the displacement of on-street parking or driving lanes, while cycle lanes, despite being expected to increase, displayed a small negative mean (-0.6%), which occurred in projects where they were merged into shared or reallocated street space. Bicycles themselves exhibited only marginal variation, with values close to zero across most projects.

Finally, outlier detection using Tukey's rule (Equation 3) further confirmed that the most extreme values appeared in cycle lanes, vegetation, sky, person and bus lane categories, highlighting that while most distributions remain stable, certain features are highly project-specific and can shift considerably depending on the intervention.

Overall, the combination of colour and segmentation analyses proved both robust and informative, providing a practical framework for systematically quantifying changes in street environments. The colour metrics effectively captured perceptual shifts in brightness, saturation and specific hues, while remaining robust to extreme variations and preserving overall colour diversity. Segmentation outputs, despite the inherent challenges of multi-class street scenes, reached competitive accuracy levels and produced meaningful indicators of spatial reallocation, highlighting consistent trends such as reductions in road space and increases in sidewalks, as well as more context-specific shifts in vegetation, sky or

pedestrian presence.

By integrating these complementary approaches, this dissertation demonstrated a feasible framework to enhance qualitative measures of street transformations, supporting higher-fidelity prediction modelling and therefore answering research question 1: *“How can emotional responses to street environments be quantified and integrated into evidence-based approaches to urban design?”*, established in chapter 1.2 Problem definition and research questions. While this approach was applied within the scope and time limit of this dissertation, it offers a foundation for further refinement and larger-scale studies.

4.2 Survey responses

The original survey was completed by 188 voluntary participants, resulting in a total of 1 504 observations due to each respondent evaluating eight interventions (Lopes, 2025). This number was reduced to 1 203 observations due to the lack of public source images found for specific interventions, as described in chapter 3.3 Image collection and preprocessing. Understanding the sample's composition is essential since these characteristics later serve as moderators in the predictive framework.

Participants represented a diverse range of 31 nationalities across 24 countries (represented in Figure 33), with the majority residing in Portugal (53%), followed by Brazil (12%) and Germany (11%). Overall, approximately 81% of participants lived in Europe, with Lisbon being the most common city of residence. This disparity in user's nationality responses suggested the reduction of categories, applied with larger scale groups.

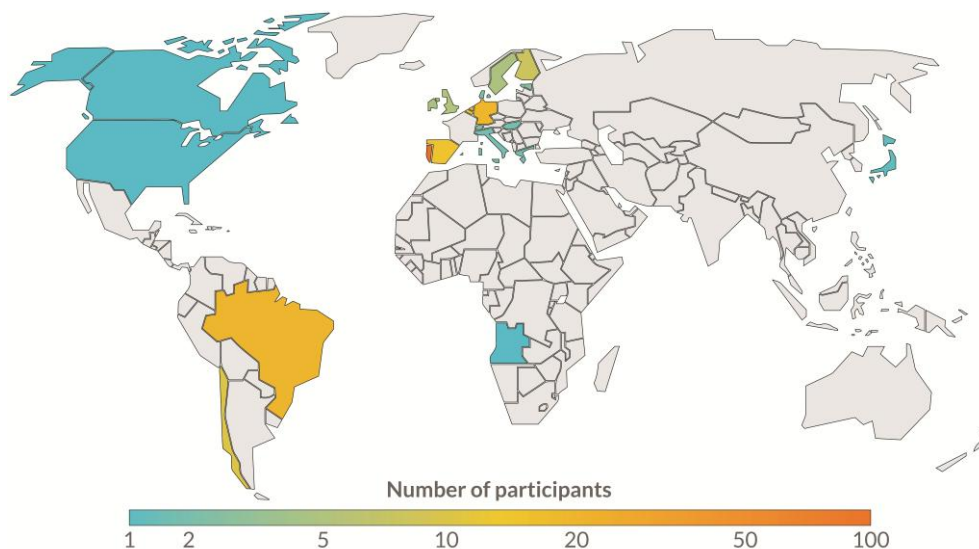


Figure 33 - Survey respondents' nationality reference.

The mean age of respondents was 35.8 years with a standard deviation of 13.2 years and a range between 20 and 85 years, reflecting a young to middle-aged adult sample. Gender distribution was slightly skewed towards females (55%), with males representing 44% and 1% identifying as non-binary or other. This again suggested the reduction in categories applied, where non-binary and other genders were excluded from the models whenever the gender variable was in use.

In terms of household characteristics, 28% of participants reported having children and around 36%

owned pets, with dogs being the most common. Vehicle ownership was high at 60%, while most respondents considered themselves urban specialists or professionals with an urban planning background (55%). Occupation-wise, most participants were employed (75%), followed by students (17%) and retired or unemployed individuals (7%). This implied the merge between the students and the retired and unemployed categories, used during the data preprocessing, due to low representation of each group separately.

Travel behaviour also varied across participants. The most common commute modes included motor vehicles (40%), walking (26%) and train or subway use (28%), with many respondents combining multiple modes for their daily commute. Commute time averaged 1.9 hours with a maximum of 4 hours and participants reported using on average two transport modes. Parking availability was mixed, with approximately 40% of respondents having free access at home, 20% paying or using a permit and 40% without access; similar patterns were observed for workplace parking.

Survey responses on the relevance of each perception per intervention used a scale from -2 to +2 for attributes such as usefulness, safety, security, stress, comfort order and liveability. While these discrete variables showed heterogeneity in responses, skewness was low, suggesting balanced distributions without extreme polarization. In contrast, preference rates were high, with 79% of interventions rated positively across the sample, which limited the SEM's capacity to learn from unpreferred observations (preference = 0).

Overall, the sample composition reflects a young, urban-aware population, due to the way it was distributed, with moderate representation across different household and commute characteristics. While the absence of monetary incentives and reliance on voluntary participation contributed to a smaller sample size, the resulting dataset provides a rich basis for exploring variations in perceptions across interventions, laying the groundwork for the predictive models that followed.

4.3 Translating visual data into perceptions

The retraining and prediction stages of the chapter 3.10 Demonstration of applied value produced a comprehensive set of statistical outputs that linked visual and spatial street characteristics with perceptual evaluations. These included six ranked coefficient plots (one per latent construct for each predictor type), exemplified in Appendix C, which illustrate the relative influence of visual and functional predictors on the perceived importance of Functionality, Habitability and Vulnerability.

For instance, within the segmentation-based results for Habitability, higher proportions of people, vegetation and sidewalks were associated with increased importance given to habitability when respondents evaluated overall preference (estimates ranging from +0.20 to +0.25). Conversely, in the colour-based framework, higher vividness and saturation values corresponded to lower importance attributed to Vulnerability (estimate = -0.12). Both tendencies align with previous literature suggesting that greenery and pedestrian infrastructure enhance comfort and liveability, while visually rich, colourful environments can reduce perceived stress or insecurity (Dabbagh, 2019; Z. Zhang et al., 2021).

The workflow also generated two ranked plots focused specifically on preference prediction, one for the colour model and another for the segmentation model, presented in Figure 34.



Figure 34 - SEM effects directional ranking (Preference on segmentation and colour systems).

These allowed the direct comparison of the relative strength and direction of predictors influencing preference. Coefficients derived from segmentation exhibited smaller magnitudes than those from colour metrics, confirming their differing sensitivity to multicollinearity, as expected from the calibration results in chapter 3.9 Predictive tool.

Besides allowing the distinction of direction per estimate on the preference prediction, coefficient ranking also enables the direct analysis of which parameters impact preference the most. The addition of cycle lane space was found to be the most effective way to increase preference in the segmentation analysis, while the percentage of pixels with ΔE_{2000} above the established threshold was the most relevant parameter in the colour system. The latter estimate revealed the survey's bias towards higher magnitude colour transformations in any direction, as long as they were noticeable. This is due to older images in the survey being less saturated, darker and noisier which meant that higher magnitude of change was positively linked with most image descriptor variables. The next strongest estimate is colour entropy which concludes that higher diversity and homogeneity of colours present in street design help increase overall preference.

Intervention-specific histograms of predicted preferences across users were then computed which helped in the quality control stage. Additionally, the tool exported a summary of the mean and standard deviation of preference probabilities and a full description file of all its calculation steps per user per intervention. The latter file enabled the management of thousands of data points per intervention into dynamic tables which allow users to compare and explore results across different combinations of demographic and habit groups.

The simpler settings from these dynamic tables, represented in Figure 33Table 4, produce the overall preference probability per intervention.

Table 4 - Colour and Segmentation overall preference predictions (Sample size: 1191).

Intervention	Preference (%)	Intervention	Preference (%)
Bus	81.72	Bus	73.84
Cycle	81.82	Cycle	87.71
MixUse	85.59	MixUse	95.20
MixUse_OP	27.15	MixUse_OP	30.16
Pedestrian	80.14	Pedestrian	87.11
Vehicle	80.01	Vehicle	67.88
<i>Colour System Prediction</i>		<i>Segmentation System Prediction</i>	

Overall preference probabilities (Table 4) indicated that the intervention prioritising mixed use urban design features received the highest predicted score, with average probabilities of 85.6% in the colour model and 95.2% in the segmentation system. Conversely, the MixUse_OP variant (the 'opposite' configuration, prioritising single-use vehicular space) displayed lower preference (27.2% to 30.2%), confirming the model's coherence between an intervention and its opposite scenario.

Interventions emphasising active modes such as Cycle or Pedestrian ranked immediately below the main MixUse scenario, agreeing that functional allocation and clear active mobility infrastructures are positively associated with habitability and overall preference. Lastly, it was possible to see that all interventions had predictions above 50%, apart from MixUse_OP. This reflects the respondent's analysed skewness towards preferring any intervention, mentioned in chapter 3.7.1 Univariate analysis.

The convergences and divergences between the two predictive systems reinforce the robustness of the framework: the different concepts in colour and segmentation metrics reflect the underlying perceptual pathways that lead to comparable preference patterns, each contributing with relevant context.

After raising the dynamic tables' complexity to allow for demographic moderator analyses, the amplitude of effects, expressed by the min-max standardised values, was used to compare group predictions. These standardised values represent the distances between demographic characteristics when evaluating street design, meaning that even if two groups have similar preferences, their differences can still be analysed.

For instance, age specific predictions highlighted a clear moderator effect in both the colour and segmentation systems (Table 5 and Table 6, respectively).

Table 5 – Preference prediction by age group in the colour system.

Intervention	Preference (%)	Standardized (0–1)	Sample
MixUse_OP	27.15		1191
Middle	23.60	0.34	496
Older	42.60	0.77	230
Young	23.29	0.33	465
Colour System Prediction			

Table 5, revealed that most interventions had stable values between age groups, apart from MixUse and its opposite scenario (MixUse_OP). In the latter transformation older respondents showed higher predicted preference scores, suggesting that visual complexity is not uniformly penalised: when saturation and colour diversity decreased, older groups interpreted them as less impactful changes when compared to younger generations. These results align with the scientific research produced by (van Leeuwen et al., 2023) which was first study to use pupillometry to show that brains become less sensitive to the intensity of colours as age progresses.

Table 6 – Preference prediction by age group in the segmentation system.

Intervention	Preference (%)	Standardized (0–1)	Sample
Bus	73.84		1191
Middle	72.39	0.41	496
Older	78.86	0.77	230
Young	72.91	0.44	465
Pedestrian	87.11		1191
Middle	87.62	0.67	496
Older	84.73	0.28	230
Young	87.75	0.69	465
Vehicle	67.88		1191
Middle	64.71	0.43	496
Older	76.18	0.79	230
Young	67.15	0.51	465
Segmentation System Prediction			

Table 6 showed that older respondents preferred the Bus and Vehicle oriented interventions when compared to young and middle-aged people (average of 0.78 to 0.45, respectively). On the contrary, younger respondents had higher prediction scores in the Pedestrian oriented approach when compared to older individuals (0.69 to 0.28, respectively). These results point to a more active younger generation while older individuals tend to use public transport or private vehicles.

Gender patterns, presented in Table 7, reinforced the interaction between chromatic cues and perceptual constructs, while finding no relevant differences in the results from the segmentation tool.

Table 7 – Preference prediction by gender in the colour system.

Intervention	Preference (%)	Standardized (0-1)	Sample
Bus	81.72		1191
Female	77.81	0.43	660
Male	86.57	0.82	531
Cycle	81.82		1191
Female	77.87	0.45	660
Male	86.72	0.82	531
Vehicle	80.01		1191
Female	75.52	0.42	660
Male	85.60	0.81	531

Colour System Prediction

Across interventions, male respondents exhibited systematically higher predicted preference (average of 0.81) than female respondents (average of 0.43). This gap reflected the larger influence of vividness and saturation on preference among men, whereas women's predictions remained more stable across interventions. This aligns with the findings in [Liu et al. \(2023\)](#) which demonstrated that males tend to prefer more “*technological and changeable*” materials while females prefer “*uniform and softer*” ones. The difference was slightly less visible in the Cycle intervention, where green hue presence increases sharply between images, indicating the relevance of the colour green in female respondent's evaluation of street design.

In terms of place of residence, differences emerged primarily in the colour system. These were represented by Table 8.

Table 8 – Preference prediction by place of residence in the colour system.

Intervention	Preference (%)	Standardized (0-1)	Sample
Cycle	81.82	0.62	1191
N/C Europe	79.77	0.53	267
South America	84.44	0.73	161
South Europe	81.89	0.62	701
MixUse	85.59	0.65	1191
N/C Europe	84.46	0.59	267
South America	87.09	0.72	161
South Europe	85.64	0.65	701

Colour System Prediction

Respondents from South American backgrounds presented higher predicted preferences for the Cycle and MixUse interventions than those from Northern/Central Europe, with standardised values around 0.73 against 0.56, respectively. These results reflected the contextual familiarity with warmer palettes and higher vegetation presence which was mimicked by the green hue of the cycle lanes in both transformations. These findings confirm that perceptual preference is not purely functional but partially shaped by visual cultural exposure and environmental reference frames. It also aligns with the research gaps identified by [Quintana \(2025\)](#) who used machine learning to demonstrate the varied perceptions of the same colour conditions globally.

Behavioural moderators were introduced in both systems to further refine how different social groups

are expected to perceive street transformations. Among these, the commuting variable produced the most distinctive contrast, seen in Table 9.

Table 9 – Preference prediction by main commute mode in the segmentation system.

Intervention	Preference (%)	Standardized (0-1)	Sample
Cycle	87.71		1191
Bicycle	90.59	0.78	131
Bus	88.18	0.65	91
Motor vehicle	88.84	0.69	480
Walking	87.75	0.63	25
Vehicle	67.88		1191
Bicycle	63.77	0.40	131
Bus	62.21	0.35	91
Motor vehicle	71.01	0.63	480
Walking	56.24	0.17	25
Segmentation System Prediction			

Respondents who regularly cycle or use public transport registered the highest preference for the Cycle intervention, with standardised values of 0.78 and 0.65, respectively. This alignment between personal mobility habits and visual preference demonstrates that recognition and identification play a crucial role in shaping preference. It shows that users are more likely to favour designs that reflect their own movement patterns.

Interestingly, even car users scored high for the same intervention (0.69) with pedestrians having the lowest score (0.63), suggesting that clearly structured cycle lanes are not necessarily perceived as competitive features but may contribute to a sense of unsafety in those whose main commuting mode is walking. In contrast, the Vehicle intervention presented a clear behavioural divide, with preferences peaking among drivers (0.63) but dropping significantly among walkers (0.17), public transport users (0.35) and cyclists (0.40). This pattern confirms the interpretive power of the segmentation framework, where functional allocation of space directly influences the perceived adequacy of each redesign according to the respondent's routine.

Family composition also showed a moderate yet meaningful influence in the segmentation-based results. Table 10 analyses the variable that defined the existence of children or dependents.

Table 10 – Preference prediction by family composition in the segmentation system.

Intervention	Preference (%)	Standardized (0-1)	Sample
Bus	73.84	0.49	1191
no children	72.96	0.44	854
has children	76.07	0.61	337
Cycle	87.71	0.63	1191
no children	88.04	0.65	854
has children	86.89	0.59	337
Vehicle	67.88	0.53	1191
no children	66.61	0.49	854
has children	71.08	0.63	337
Segmentation System Prediction			

Households with children expressed higher predicted preferences particularly for the Bus and Vehicle configurations, reaching average standardised values of 0.62. This reinforces the notion that the

presence of dependents heightens the salience of safety and comfort cues in both the mobility habits and the street environment. Even though the cycle-oriented scenario remained positively evaluated, the preference gap between having children or not corroborated that safety related features such as wider sidewalks, slower traffic lanes or clear spatial separation are more highly valued when personal responsibility for others is involved.

Lastly, the combination of two moderators such as car ownership and parking access at work introduced a meaningful layer to the possible prediction analyses in Table 11. The segmentation-based predictions showed that having a private vehicle does not necessarily correspond to stronger support for car-oriented designs. Rather, respondents with car access displayed higher preference for the Pedestrian intervention (86.3%) compared to the Vehicle one (71.1%).

Table 11 – Preference prediction by family composition in the segmentation system.

Intervention	Preference (%)	Standardized (0-1)	Sample
Pedestrian	87.11		1191
has car	86.33		721
Free	86.02	0.45	399
Paid/Permit	86.72	0.55	322
Vehicle	67.88		1191
has car	71.09		721
Free	71.70	0.65	399
Paid/Permit	70.34	0.61	322

Segmentation System Prediction

Within this subgroup, respondents with Paid/Permit access to parking at work exhibited the highest stability across interventions, reflecting a pragmatic rather than ideological stance towards mobility infrastructure. This group's moderate but consistent approval across scenarios implies that improved walkability and regulated parking are not perceived as conflicting goals. On the other hand, individuals with free parking access at work displayed lower preference for the Pedestrian intervention while showing higher preference scores in the Vehicle-oriented transformation, when compared to the average rates.

Latent constructs such as Functionality, Habitability and Vulnerability captured how individuals interpret streets, with visual descriptors such as saturation, hue diversity and spatial allocation, contributing to perceived quality. This demonstrates that psychological evaluations can be inferred directly from imagery without relying on self-reports.

Methodologically, segmentation captured spatial and functional information (e.g., vehicles, sidewalks, greenery), while colour metrics captured affective properties (e.g., contrast, chromatic balance). Both systems identified MixUse and active-mobility scenarios as preferred, with colour highlighting visual balance and segmentation highlighting spatial functionality. The convergent rankings confirmed construct validity: interventions enhancing pedestrian and mixed-use qualities systematically increased predicted preference, with Functionality and Habitability positively and Vulnerability negatively, affecting perceptions. Moderator effects further demonstrated that sustainable designs are widely preferred but interpreted through distinct perceptual filters.

Addressing question 2: *“Can exploratory predictive tools anticipate public preference to street redesigns before implementation?”*, moderator analyses showed coherent variations across age,

gender, mobility habits, family composition and regional background, among others. Older respondents preferred calmer, less complex streets, younger participants valued active-mobility configurations and gender, region and behaviour influenced sensitivity to colour, vegetation and space allocation.

The results obtained in this study aligned with several of the relationships identified by [Lopes \(2025\)](#). In both analyses, the visual presence of vegetation and pedestrians was associated with higher preference whereas vehicle presence and car-oriented layouts reduced it. These consistencies indicated that the perceptual importance of greenery, human activity and balanced spatial allocation persisted across methodological approaches. Both studies also found that individual mobility habits and family composition shaped how participants evaluated interventions, confirming the behavioural relevance of lived mobility experience in perceptual assessments.

While [Lopes \(2025\)](#) captured perceptual dynamics using qualitative image descriptors and logistic regression, the present framework extended this by translating constructs into quantitative visual and chromatic predictors. The perceptual dimensions themselves remained consistent with the previous work, with the main differences being in the measurement, analytical and predictive tool approach. Unlike the logistic regression, the structural equation modelling allowed for mediation and moderation effects and supported a term by term adaptable framework that could identify which factors were most relevant to the community in question. This approach thus strengthened the analytical connection between visual evidence and public perception while complementing, rather than substituting, the rich insights obtainable from qualitative descriptors.

Overall, the tool reproduced realistic, interpretable preference patterns and could serve as a decision-support instrument, allowing planners to iterate and anticipate social reactions before implementation or investment. By embedding automated moderator effects, the framework preserved perceptual diversity, advancing an inclusive and adaptable evidence-based approach to urban design.

5 Limitations and future work

This final chapter reflects on the methodological, practical and theoretical dimensions of the research, highlighting its implications, limitations and future directions. While the primary aim of this dissertation was the conception and validation of a predictive tool linking visual data and subjective perceptions for urban design applications, it also sought to critically examine the challenges inherent to such integration.

The following sections assess the framework's operational constraints, data dependencies and representational assumptions, as well as its potential to inform more participatory and evidence-based planning practices. By situating the developed tool within broader methodological and technological frontiers, this chapter identifies both the current boundaries of perception modelling and the pathways through which future research could extend its scope, robustness and social relevance.

5.1 Implications and limitations

The second research question: "Can exploratory predictive tools anticipate public preference to street redesigns before implementation?", highlighted the practical and methodological scope of this dissertation. By combining corrected semantic segmentation, perceptual colour metrics, univariate/bivariate analysis, polychoric EFA/CFA approaches and SEM-based modelling, the thesis operationalised a predictive tool capable of simulating public reactions to urban interventions. This workflow was demonstrated through five alternative designs for a single street in chapter 4.3 Translating visual data into perceptions. The study aimed to bridge the gap between advanced visual analytics and participatory planning, demonstrating that street interventions could be evaluated not only through self-reports but also through objective, image-derived features.

From a practical standpoint, the framework offered stakeholders (urban planners, architects and policymakers) a decision-support instrument that translated complex visual and functional features of streets into interpretable constructs, including the Importance of Functionality, Habitability and Vulnerability when evaluating street design. These constructs allow the anticipation of population and subgroup specific preferences prior to the allocation of time and resources.

Colour and segmentation predictors captured complementary aspects of the urban experience, representing affective visual qualities and spatial functionality, respectively. This dual perspective provided actionable guidance for identifying interventions that were likely to increase public preference. The approach aligned with calls in recent literature, including by [Lopes \(2025\)](#), for tools that integrated perception research into planning while preserving transparency, interpretability and the capacity to account for demographic heterogeneity.

However, the results revealed several limitations associated with its dataset. The preference distribution was positively skewed, with an overall rating of 78.55%, which biased predicted probabilities toward higher values and compressed absolute differences across subgroups. This effect reduced the model's sensitivity to less-preferred interventions, though standardisation using min-max scores allowed meaningful assessment of relative deviations. Furthermore, the dataset was limited to specific street types and contexts, reducing the generalisability of the results to streets with distinct morphological

characteristics or different urban cultures.

Survey related limitations also shaped the outcomes. Respondent demographics were skewed toward younger, digitally literate participants, which limited the representativeness of the example scenario findings. Although demographic moderators were explicitly incorporated into the SEM framework, older, less technologically engaged or socioeconomically disadvantaged groups were underrepresented. This limitation suggested that while the tool captured broad perceptual tendencies, caution was required when generalising findings to the entire population.

Image and colour measures introduced additional constraints. Seasonal, lighting and time-of-day variations affected colour metrics, complicating comparisons across images. Colour descriptors, including perceptual brightness, saturation, hue diversity and ΔE_{2000} distances, were sensitive to such variability, requiring careful standardisation. Semantic segmentation models, even after manual correction, retained biases from training datasets, occasionally misclassifying features that were critical to perception, such as street furniture or signage. Moreover, to comply with copyright restrictions on Google Street View imagery, intervention photographs were sourced from multiple public repositories. While this ensured ethical compliance, it introduced further variability in perspective, scale and image quality, requiring deliberate correction and validation procedures.

Methodological and statistical limitations were also present. Polychoric EFA/CFA approaches assumed independence of observations and SEM relied on large-sample stability, linearity of relationships and correctly specified factor structures. While the SEM framework enabled automated moderation testing and the operationalisation of latent constructs, these assumptions constrained model flexibility and limited its ability to capture highly non-linear or complex interactions between visual and perceptual variables. The predictive tool generated robust relative rankings but should be interpreted as prioritisation guidance rather than precise forecasts.

Lastly, the colour and segmentation systems are treated with the same weights in the final predictions. Future work should explore the relationship between segmentation and colour variation across different cultures and cities and estimate the relative weight of importance given to each group of predictors, when evaluating urban intervention preference.

Despite these constraints, the thesis addressed key research gaps identified in the literature. Previous studies frequently relied on static images, single-city datasets or unrelated image comparisons and often excluded colour metrics. They rarely captured before and after intervention changes or integrated demographic moderators in an exploratory manner. By combining before and after imagery, corrected semantic segmentation, perceptual colour metrics and survey-derived demographic and mobility information, this research quantified perceptual changes attributable to specific urban interventions. Latent constructs were operationalised through SEM and translated into a predictive tool that was both interpretable and transferable, demonstrating how visual data could meaningfully inform automatic evidence-based participatory planning in urban design.

In conclusion, the dissertation demonstrated that combining visual data with subjective ratings could meaningfully inform design choices, preserve perceptual diversity and support inclusive planning processes. It established a replicable methodology and practical decision-support prototype, highlighting both the opportunities of integrating subjective and visual data and the limitations inherent to

contemporary datasets, models and survey practices. The study advanced the capacity for urban planners, designers and policymakers to anticipate public reactions, supporting more transparent, participatory and evidence-based approaches to street design.

5.2 Future directions

While this dissertation contributed a novel workflow, predictive framework and practical tool for integrating visual data and subjective perception into urban planning, it represented only an initial step toward fully automated, evidence-based and participatory design. Future research could build on these foundations to extend methodological sophistication, improve generalisability and enhance the operational value of perception modelling.

One avenue for advancement would involve refining semantic segmentation through hybrid approaches that combined pixel-wise classification with instance-level or object-based methods, while keeping the model light enough to run on any workstation. Such approaches could reduce boundary errors and more accurately represent complex urban features, including trees of different species, small-scale street furniture, signage, paving materials and façade textures, which could strongly influence perceived safety, comfort and aesthetic evaluation. Weakly supervised or self-supervised learning techniques could also mitigate the dependency on costly pixel-level annotations, allowing the creation of larger, more diverse datasets spanning multiple cities, climates and street typologies. Besides, expanding segmentation outputs remains a key goal, especially for distinguishing elements that contribute to affective responses, such as vegetation density, surface reflectivity or material.

Expanding the diversity and modality of training datasets could be another critical research direction. The integration of three-dimensional point clouds, photogrammetric reconstructions or LiDAR scans could capture volumetric relationships, occlusions and spatial configurations that static images alone could not convey. Such datasets would enable interactive video sequences, 3D rendering comparisons or immersive simulations of street interventions, allowing planners, designers and citizens to explore alternative designs dynamically. This would make it possible to evaluate both functional and aesthetic trade-offs, such as the effect of sidewalk widening on perceived safety or how tree placement modulates spatial comfort.

Additionally longitudinal data collection could reveal how public perception evolved over time, capturing both immediate reactions and longer-term behavioural or attitudinal shifts, which is especially important for climate-adaptive or phased interventions.

Generative modelling presents another promising frontier. These could simulate “after-intervention” scenarios from existing street imagery, enabling planners and stakeholders to explore “what-if” design alternatives before implementation, as exemplified in chapter 4.3 Translating visual data into perceptions. For instance, generative simulations could explore the impact of adding street trees, reallocating parking space to cycle lanes or altering façade materials, providing visual and quantitative estimates of perceptual changes. Besides, integrating segmentation outputs, perceptual colour metrics and latent importance measures within generative pipelines could ensure that simulated interventions remained interpretable and perceptually grounded. However, to maintain reliability, future studies would need to address challenges related to semantic consistency, lighting, vegetation type and accurate

representation of fine urban details, preventing visual artefacts from misleading public interpretation.

Future research could also integrate multimodal perception measures to enrich survey-based assessments. Physiological and behavioural indicators, including eye tracking, pedestrian movement patterns, heart rate variability or stress-related biomarkers, could triangulate self-reported preferences and strengthen model reliability. This approach would allow models to account for unconscious or automatic responses to street design, complementing subjective ratings and improving predictive accuracy. Furthermore, error quantification could be enhanced by establishing confidence intervals around pixel-change estimates, reliability measures during correction phases or sensitivity analyses across demographic subgroups, increasing transparency and robustness.

Practical and participatory extensions of the methodology could transform perception modelling into a proactive, anticipatory instrument. Immersive AR environments could allow citizens and decision-makers to interactively explore multiple design scenarios, evaluating spatial layouts, visual composition and functional allocations. Semi-automated correction workflows or active learning loops could reduce manual bias, allowing models to improve iteratively as new interventions and perception data became available. By integrating interactive simulations with generative modelling, explainable AI and human-perceived importance metrics, perception models could support anticipatory planning rather than merely evaluating past or current conditions.

Additionally, climate-sensitive and equity-focused applications could simulate scenarios such as flood-resilient streetscapes, pedestrian-first interventions or enhanced tree canopy coverage, providing stakeholders with actionable insights that balanced aesthetics, functionality and sustainability.

Ethical considerations should remain central in future developments. The generation of synthetic images, simulations or immersive experiences must preserve semantic consistency and realism to avoid misleading stakeholders. Transparent documentation of assumptions, correction protocols and model limitations would ensure responsible communication of predictions, particularly when influencing public expectations. By combining robust, reproducible methodologies with ethical safeguards, future work could operationalise urban perception analysis as a proactive, transparent and socially responsible instrument, enabling planners and communities to explore “what might be” in addition to understanding “what was” or “what is.”

In summary, building on the workflow developed in this dissertation, future research could integrate semantic segmentation, perceptual colour metrics, generative modelling, explainability methods, multimodal perception measures, interactive 3D, VR and AR simulations and longitudinal tracking. These extensions would transform urban perception modelling from a descriptive or reactive tool into a proactive, anticipatory instrument, capable of simulating potential outcomes, guiding design decisions and supporting participatory, evidence-based urban planning across diverse contexts and temporal horizons. By operationalising perception analysis as both a predictive and explanatory framework, subsequent studies could provide planners, designers and stakeholders with actionable insights that balanced aesthetics, functionality, inclusivity and sustainability in future urban interventions.

5.3 Final remarks

This dissertation set out to develop and validate a predictive tool capable of linking visual

transformations in street environments with subjective perceptions of urban quality. In doing so, it bridged computational image analysis and behavioural research, proposing a workflow that could translate objective, image-derived indicators into interpretable perceptual constructs. Rather than aiming to produce conventional untransferable empirical findings, the results demonstrated how automated yet transparent analytical processes could support participatory, anticipatory and evidence-based urban design. The tool itself thus represents the central outcome of this research: an operational prototype through which perception can be quantified, visualised and meaningfully integrated into decision-making.

Through the combination of semantic segmentation, perceptual colour metrics, structural equation modelling and exploratory demographic moderation, the dissertation advanced an interdisciplinary methodology that unites visual computing, environmental psychology and urban design. Its contribution lies in proving that image-based representations can serve not only descriptive or aesthetic purposes but also predictive and interpretative ones. By enabling the simulation of perceptual responses before an intervention occurs, the framework opens new opportunities for dialogue between planners, designers and citizens, helping anticipate public reactions and design more inclusive, visually coherent streetscapes.

At a broader level, the research corroborated that perception is not an abstract or purely subjective phenomenon but a measurable and reproducible aspect of the urban experience. Integrating it into computational models allows planners to consider lived experience alongside functional efficiency and sustainability, aligning design decisions with how environments are used, seen and felt by their users. The thesis thus contributes to a growing paradigm of human-centred urban analytics, where the visual and affective dimensions of the city become quantifiable, discussable and actionable within planning workflows.

While the framework remains a first iteration, its methodological coherence, adaptability and interpretability position it as a foundation for future research and practice. The proposed extensions from generative modelling to multimodal perception in interactive 3D and AR environments illustrate a broader vision of perception modelling as a continuous, participatory process rather than a static evaluation. The dissertation underscores the value of linking data-driven analysis with human experience, reaffirming that the future of urban design lies not only in understanding how cities function but also in understanding how they are perceived.

References

- Alcoforado, M. J., Andrade, H., Oliveira, S., Festas, M. J., & Rosa, F. (2009). Alterações climáticas e desenvolvimento urbano. In DGOTDU - Direção-Geral do Ordenamento do Território e Desenvolvimento Urbano (Ed.), *Política de Cidades* (Política de Cidades, Vol. 4).
- Alemaskin, K., Zloczower, I. M., & Kaufman, M. (2005). Entropic Analysis of Color Homogeneity. *Physics Faculty Publications*, 234. https://engagedscholarship.csuohio.edu/sciphysics_facpub
- Alves, M. (2025, October 31). *U-Shift - Modelling Preferences on Street Redesigns*. Github. <https://github.com/U-Shift/Modelling-Preferences-on-Street-Redesigns>
- Arellana, J., Saltaín, M., Larrañaga, A. M., Alvarez, V., & Henao, C. A. (2019). Urban walkability considering pedestrians' perceptions of the built environment: a 10-year review and a case study in a medium-sized city in Latin America. *Transport Reviews*, 40(2), 183–203. <https://doi.org/10.1080/01441647.2019.1703842>
- Arnett, J. J. (2000). Emerging adulthood: A theory of development from the late teens through the twenties. *American Psychologist*, 55(5), 469–480. <https://doi.org/https://doi.org/10.1037/0003-066X.55.5.469>
- Atif, R. M. D. (2024, October 6). *Z-Score Normalization: Standardizing Your Data for Better Model Performance*. Medium. <https://medium.com/@raodayanatif/z-score-normalization-standardizing-your-data-for-better-model-performance-43102cc09ed0>
- Atiquel Haq, S. M., Islam, M. N., Siddhanta, A., Ahmed, K. J., & Chowdhury, M. T. A. (2021). Public Perceptions of Urban Green Spaces: Convergences and Divergences. In *Frontiers in Sustainable Cities* (Vol. 3). Frontiers Media S.A. <https://doi.org/10.3389/frsc.2021.755313>
- Badrinarayanan, V., Handa, A., & Cipolla, R. (2015). *SegNet: A Deep Convolutional Encoder-Decoder Architecture for Robust Semantic Pixel-Wise Labelling*. <http://arxiv.org/abs/1505.07293>
- Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2), 88–97.
- Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research Methodology in the Social Sciences* (T. D. Little, Ed.; 2nd Edition). The Guilford Press. www.guilford.com/MSS
- Business Analytics Institute. (2025, September 24). *Using Mixed Precision and Gradient Accumulation*.
- Canova, S., Cortinovis, D. L., & Ambrogi, F. (2017). How to describe univariate data. *Journal of Thoracic Disease*, 9, 1741–1743. <https://doi.org/10.21037/jtd.2017.05.80>
- Cavaco, C., Vilares, E., Rosa, F., Magalhães, M., & Esteves, N. (2015). *Sustainable Cities 2020* (Directorate-General for Territory, Ed.). Ministry of Environment, Spatial Planning and Energy. www.dgterritorio.pt
- Chen, C., Li, H., Luo, W., Xie, J., Yao, J., Wu, L., & Xia, Y. (2022). Predicting the effect of street environment on residents' mood states in large urban areas using machine learning and street view images. *Science of The Total Environment*, 816. <https://doi.org/10.1016/j.scitotenv.2021.151605>
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). *DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs*. <http://arxiv.org/abs/1606.00915>
- Chen, L.-C., Papandreou, G., Schroff, F., & Adam, H. (2017). *Rethinking Atrous Convolution for Semantic Image Segmentation*. <http://arxiv.org/abs/1706.05587>
- Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*. <http://arxiv.org/abs/1802.02611>
- Chen, Tan, V., León-Pérez, G., Honnold, J., & Aytar, V. (2024). *The Craft of Sociological Research: Principles and Methods of Collecting, Analyzing, and Presenting Social Science Data*. Fairfax: Virtual Library of Virginia.
- Christopher Westland, J. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, 9(6), 476–487. <https://doi.org/10.1016/j.elerap.2010.07.003>
- Cityscapes Dataset. (2025, September 24). *Dataset Overview*.
- Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). *The Cityscapes Dataset for Semantic Urban Scene Understanding*. <http://arxiv.org/abs/1604.01685>
- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Research, and Evaluation Practical Assessment, Research, and Evaluation*, 10(7).
- Crouse, D. W. (2004). Traffic Calming: A Social Issue. *Bulletin of Science, Technology and Society*,

- 24(2), 138–144. <https://doi.org/10.1177/0270467604263564>
- Cui, Q., Zhang, Y., Yang, G., Huang, Y., & Chen, Y. (2023). Analysing gender differences in the perceived safety from street view imagery. *International Journal of Applied Earth Observation and Geoinformation*, 124, 103537. <https://doi.org/10.1016/j.jag.2023.103537>
- Dabbagh, E. (2019). The Effects of Color and Light on the Beautification of Urban Space and the Subjective Perception of Citizens. *IJESI - International Journal of Engineering Science Invention*, 8(3), 20–25. www.ijesi.org
- De Nadai, M., Vieriu, R. L., Zen, G., Dragicevic, S., Naik, N., Caraviello, M., Hidalgo, C. A., Sebe, N., & Lepri, B. (2016). *Are Safer Looking Neighborhoods More Lively? A Multimodal Investigation into Urban Life*. <http://arxiv.org/abs/1608.00462>
- de Winter, J. C. F., Gosling, S. D., & Potter, J. (2016). Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological Methods*, 21, 273–290. <https://doi.org/10.1037/met0000079>
- Dive into Deep Learning. (2025, September 24). 3.7. *Weight Decay*.
- Dong, G., Yan, Y., Shen, C., & Wang, H. (2020). Real-Time High-Performance Semantic Image Segmentation of Urban Street Scenes. *Journal of Latex Class Files*. <http://arxiv.org/abs/2003.08736>
- Dubey, A., Naik, N., Parikh, D., Raskar, R., & Hidalgo, C. A. (2016). *Deep Learning the City : Quantifying Urban Perception At A Global Scale*. <http://arxiv.org/abs/1608.01769>
- Duncan, S., Chachra, A., Prakash, A., & Prati, F. (2016). *Guia Global de Desenho de Ruas*. <https://globaldesigningcities.org/publication/global-street-design-guide/>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the Use of Exploratory Factor Analysis in Psychological Research. *Psychological Methods*, 4(3), 272–299.
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). *A Review on Deep Learning Techniques Applied to Semantic Segmentation*. <http://arxiv.org/abs/1704.06857>
- Gehl, J. (1936). *Cities For People*.
- Geldhof, G. J., Pornprasertmanit, S., Schoemann, A. M., & Little, T. D. (2013). Orthogonalizing Through Residual Centering. *Educational and Psychological Measurement*, 73(1), 27–46. <https://doi.org/10.1177/0013164412445473>
- González Mendoza, B., Acioly, C., Zaltman, G., Maia, G., Santos, H., Tuszel, L., & Jardim, L. (2019). *Consulta Cidades Sustentáveis* (Colab & ONU-Habitat, Eds.). www.unhabitat.org
- Google. (2025, September 24). *Geo Guidelines - Google Maps, Google Earth, and Street View*.
- Gorzaldini, M. N. (2016). The Effects of Colors on the Quality of Urban Appearance. *Mediterranean Journal of Social Sciences*, 7(5). <https://doi.org/10.5901/mjss.2016.v7n5p225>
- Gössling, S. (2020). Why cities need to take road space from cars - and how this could be done. *Journal of Urban Design*, 25(4), 443–448. <https://doi.org/10.1080/13574809.2020.1727318>
- Gurita, A., & Mocanu, I. G. (2021). Image segmentation using encoder-decoder with deformable convolutions. *Sensors*, 21(1570), 1–27. <https://doi.org/10.3390/s21051570>
- Habekost, M. (2013). Which color differencing equation should be used? *International Circular of Graphic Education and Research*, 6. <http://www.idealliance.org/downloads/idealliance-2006-swop->
- Hair JR., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (n.d.). *Multivariate Data Analysis*.
- Hao, N., Li, X., Han, D., & Nie, W. (2024). Quantifying the Impact of Street Greening during Full-Leaf Seasons on Emotional Perception: Guidelines for Resident Well-Being. *Forests*, 15. <https://doi.org/10.3390/f15010119>
- Hartigan, J. A., & Kleiner, B. (1981). Mosaics for Contingency Tables. *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*, 268–269.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition*. <http://arxiv.org/abs/1512.03385>
- He, N., & Li, G. (2021). Urban neighbourhood environment assessment based on street view image processing: A review of research trends. *Environmental Challenges*, 4. <https://doi.org/10.1016/j.envc.2021.100090>
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2008). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality and Quantity*, 44, 153–166. <https://doi.org/10.1007/s11135-008-9190-y>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hussein, H., Nevill, C. R., Meffen, A., Abrams, K. R., Bujkiewicz, S., Sutton, A. J., & Gray, L. J. (2022).

- Double-counting of populations in evidence synthesis in public health: a call for awareness and future methodological development. *BMC Public Health*, 22(1827). <https://doi.org/10.1186/s12889-022-14213-6>
- Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., Zhao, T., Ma, R., Fujiwara, K., Ouyang, J., Quintana, M., & Biljecki, F. (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. <https://doi.org/10.1016/j.compenvurbsys.2025.102283>
- Jaglarz, A. (2023). Perception of Color in Architecture and Urban Space. *Buildings*, 13(2000). <https://doi.org/10.3390/buildings13082000>
- Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika*, 39(1), 31–36.
- Kavengik. (2025, January 31). *Orthogonal rotation: Enhancing interpretability of principal components using the varimax technique*. Medium. <https://medium.com/@kavengik/orthogonal-rotation-enhancing-interpretability-of-principal-components-using-the-varimax-technique-5d64d5f51301>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (T. D. Little, Ed.; Fourth Edition). The Guilford Press.
- Kumar Bhunia, A., Bhattacharyya, A., Banerjee, P., Roy, P. P., & Murala, S. (n.d.). *A Novel Feature Descriptor for Image Retrieval by Combining Modified Color Histogram and Diagonally Symmetric Co-occurrence Texture Pattern*.
- Larkin, A., Gu, X., Chen, L., & Hystad, P. (2021). Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landscape and Urban Planning*, 216. <https://doi.org/10.1016/j.landurbplan.2021.104257>
- Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936–949. <https://doi.org/10.3758/s13428-015-0619-7>
- Li, T., Jiang, C., Bian, Z., Wang, M., & Niu, X. (2020). Semantic Segmentation of Urban Street Scene Based on Convolutional Neural Network. *Journal of Physics: Conference Series*, 1682(1). <https://doi.org/10.1088/1742-6596/1682/1/012077>
- Li, X., Zhang, C., & Li, W. (2015). Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset. *ISPRS International Journal of Geo-Information*, 4, 1166–1183. <https://doi.org/10.3390/ijgi4031166>
- Lightly. (2025, September 24). *Cross-entropy Loss*.
- Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods*, 24, 452–467. <https://doi.org/10.1037/met0000230>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). *Microsoft COCO: Common Objects in Context*. <http://arxiv.org/abs/1405.0312>
- Liu, X., Zhu, L., Zhu, S., & Luo, L. (2022). *Distance Guided Channel Weighting for Semantic Segmentation*. <http://arxiv.org/abs/2004.12679>
- Liu, Y., Zhou, Z., & Xu, Y. (2023). Design Element Preferences in Public Facilities: An Eye Tracking Study. *Land*, 12(1411). <https://doi.org/10.3390/land12071411>
- Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully Convolutional Networks for Semantic Segmentation*. <http://arxiv.org/abs/1411.4038>
- Lopes, M. I. (2025). *Perceptions of Urban Change: Exploring Before-and-After Visual Preferences for Mobility*.
- Loshchilov, I., & Hutter, F. (2019). *Decoupled Weight Decay Regularization*. <http://arxiv.org/abs/1711.05101>
- Loyola, M., Nelson, J. D., Clifton, G., & Levinson, D. (2023). The influence of cycle lanes on road users' perception of road space. *Urban, Planning and Transport Research*, 11(1). <https://doi.org/10.1080/21650020.2023.2195894>
- Luo, M. R., Cui, G., & Rigg, B. (2001). The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application*, 26(5), 340–350. <https://doi.org/10.1002/col.1049>
- Ma, Q., Wei, X., & Ren, G. (2024). Street Design Guidelines and Urban Road System Improvement: From Traffic Volume to Space Quality. *Transportation Research Procedia*, 82, 3266–3282. <https://doi.org/10.1016/j.trpro.2024.12.095>
- Marôco, J. (2014). *Análise de Equações Estruturais - Fundamentos teóricos, Software & Aplicações* (Report Number - Análise e Gestão de Informação, Ed.; 2ª Edição).
- Marshall, E., Boggis, E., Patel, C., Emmett, M., Owen, A., Russel, J., & Fieller, N. (n.d.). *The Statistics Tutor's Quick Guide to Commonly Used Statistical Tests*.
- Merlo, L., Johansson, C., Nilson, F., & Chapman, D. (2025). State of the art for walking as a transport

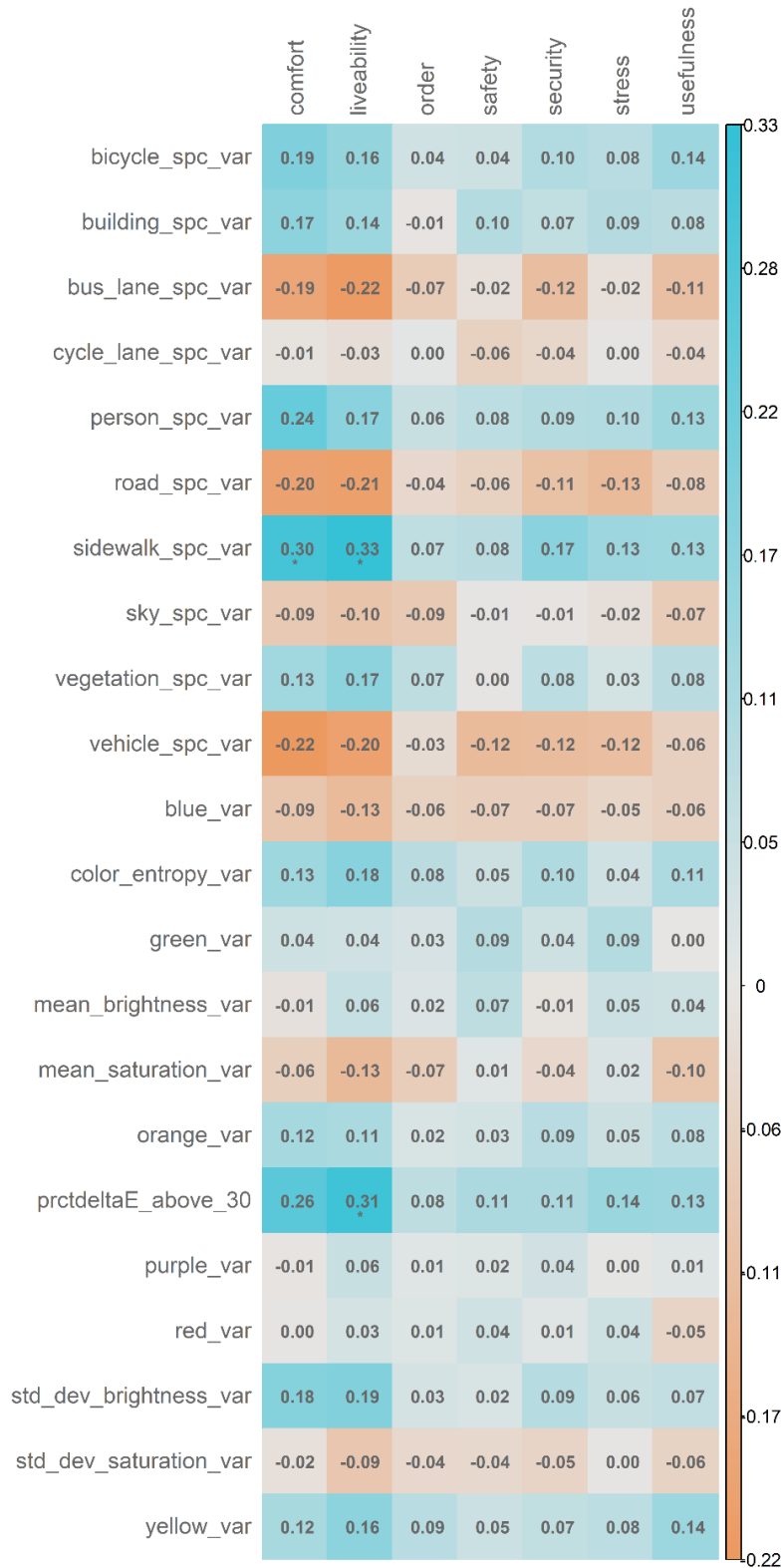
- mode within 15-minute cities. In *Urban, Planning and Transport Research* (Vol. 13, Issue 1). Routledge Taylor and Francis. <https://doi.org/10.1080/21650020.2025.2456186>
- Mindspore. (2025, September 24). *Gradient Accumulation Algorithm*.
- Moreno-Vera, F., Lavi, B., & Poco, J. (2021). Urban Perception: Can we understand why a street is safe? In *Advances in Computational Intelligence* (Vol. 13067). https://doi.org/10.1007/978-3-030-89817-5_21
- Muller, E., Gemmell, E., Choudhury, I., Nathvani, R., Metzler, A. B., Bennett, J., Denton, E., Flaxman, S., & Ezzati, M. (2022). *City-Wide Perceptions of Neighbourhood Quality using Street View Images*. <http://arxiv.org/abs/2211.12139>
- National University. (2025, July 31). *Statistics Resources - Point Biserial*. Academic Success Center. <https://resources.nu.edu/statsresources>
- Neuhold, G., Ollmann, T., Rotabui, S., & Kontschieder, P. (n.d.). *The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes*. www.mapillary.com
- Park, Y., & Garcia, M. (2019). Pedestrian safety perception and urban street settings. *International Journal of Sustainable Transportation*, 14(11), 860–871. <https://doi.org/10.1080/15568318.2019.1641577>
- Paul, M., Luc, C. M., Gool, V., & Timofte, R. (2020). *Efficient Video Semantic Segmentation with Labels Propagation and Refinement*. 2873–2882. <https://confluence.uniqfeed.com/plugins/drawio/>
- Pereira Martins, I. (2021). *Como é que as cidades europeias se estão a adaptar às alterações climáticas e a preparar para um futuro sustentável?* https://www.eea.europa.eu/ds_resolueuid/4a207d8ff93743b2afae57d2e6cfe6c6
- PyTorch - torch.utils.data. (2025, September 24). *Single- and Multi-process Data Loading*.
- Quintana, M., Gu, Y., Liang, X., Hou, Y., Ito, K., Zhu, Y., Abdelrahman, M., & Biljecki, F. (2025). *Global urban visual perception varies across demographics and personalities*. <http://arxiv.org/abs/2505.12758>
- Rajapaksa, D., Islam, M., & Managi, S. (2018). Pro-environmental behavior: The role of public perception in infrastructure and the social factors for sustainable development. *Sustainability*, 10(4). <https://doi.org/10.3390/su10040937>
- Rashid, S. A., Wahab, M. H., Rani, W. N. M. W. M., & Ismail, S. (2017). Safety of street: The role of street design. *AIP Conference Proceedings*, 1891. <https://doi.org/10.1063/1.5005341>
- Reza, A., Manley, E., Chen, S., Chaudhary, S., & Elafros, J. (n.d.). *SegBuilder: A Semi-automatic Annotation Tool for Segmentation*.
- Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods Under Suboptimal Conditions. *Psychological Methods*, 17(3), 354–373. <https://doi.org/10.1037/a0029315.supp>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. <http://arxiv.org/abs/1505.04597>
- Sádaba, J., Alonso, Y., Latasa, I., & Luzarraga, A. (2024). Towards Resilient and Inclusive Cities: A Framework for Sustainable Street-Level Urban Design. *Urban Science*, 8. <https://doi.org/10.3390/urbansci8040264>
- Sharma, G., Wu, W., & Dalal, E. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1), 21–30. <https://doi.org/10.1002/col.20070>
- Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(60). <https://doi.org/10.1186/s40537-019-0197-0>
- Smith, L. N. (2018). *A disciplined approach to neural network hyper-parameters: Part 1 -- learning rate, batch size, momentum, and weight decay*. <http://arxiv.org/abs/1803.09820>
- Smith, L. N., & Topin, N. (2018). *Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates*. <http://arxiv.org/abs/1708.07120>
- Smyth, R., & Johnson, A. (2025). *Factor Analysis*. <https://www.uwo.ca/fhs/tc>
- Tang, J., & Long, Y. (2019). Measuring visual quality of street space and its temporal variation: Methodology and its application in the Hutong area in Beijing. *Landscape and Urban Planning*, 191. <https://doi.org/10.1016/j.landurbplan.2018.09.015>
- University of Regina. (n.d.). *Measures of Association* (pp. 767–786).
- Valença, G., Azevedo, C., Moura, F., & Morais de Sá, A. (2025). Creating visualizations using generative AI to guide decision-making in street designs: A viewpoint. In *Journal of Urban Mobility* (Vol. 7). Elsevier Ltd. <https://doi.org/10.1016/j.urbmob.2025.100104>
- van Laarhoven, T. (2017). *L2 Regularization versus Batch and Weight Normalization*. <http://arxiv.org/abs/1706.05350>

- van Leeuwen, J. E. P., McDougall, A., Mylonas, D., Suárez-González, A., Crutch, S. J., & Warren, J. D. (2023). Pupil responses to colorfulness are selectively reduced in healthy older adults. *Scientific Reports*, 13. <https://doi.org/10.1038/s41598-023-48513-7>
- Verma, D., Jana, A., & Ramamritham, K. (2020). Predicting human perception of the urban environment in a spatiotemporal urban setting using locally acquired street view images and audio clips. *Building and Environment*, 186. <https://doi.org/10.1016/j.buildenv.2020.107340>
- Webster, D. C. (1998). *Traffic calming-Public attitude studies: a literature review*.
- Wu, Y., Liu, Q., Hang, T., Yang, Y., Wang, Y., & Cao, L. (2024). Integrating restorative perception into urban street planning: A framework using street view images, deep learning, and space syntax. *Cities*, 147. <https://doi.org/10.1016/j.cities.2024.104791>
- Yao, Y., Liang, Z., Yuan, Z., Liu, P., Bie, Y., Zhang, J., Wang, R., Wang, J., & Guan, Q. (2019). A human-machine adversarial scoring framework for urban perception assessment using street-view images. *International Journal of Geographical Information Science*, 33(12), 2363–2384. <https://doi.org/10.1080/13658816.2019.1643024>
- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?* <http://arxiv.org/abs/1411.1792>
- Yu, M., Zheng, X., Qin, P., Cui, W., & Ji, Q. (2024). Urban Color Perception and Sentiment Analysis Based on Deep Learning and Street View Big Data. *Applied Sciences*, 14. <https://doi.org/10.3390/app14209521>
- Zhang, Y., Wang, M., Li, J., Chang, J., & Lu, H. (2022). Do Greener Urban Streets Provide Better Emotional Experiences? An Experimental Study on Chinese Tourists. *International Journal of Environmental Research and Public Health*, 19(19). <https://doi.org/10.3390/ijerph192416918>
- Zhang, Z., Zhuo, K., Wei, W., Li, F., Yin, J., & Xu, L. (2021). Emotional responses to the visual patterns of urban streets: Evidence from physiological and subjective indicators. *International Journal of Environmental Research and Public Health*, 18(18). <https://doi.org/10.3390/ijerph18189677>
- Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). *Pyramid Scene Parsing Network*. <http://arxiv.org/abs/1612.01105>
- Zhao, W., Tan, L., Niu, S., & Qing, L. (2024). Assessing the Impact of Street Visual Environment on the Emotional Well-Being of Young Adults through Physiological Feedback and Deep Learning Technologies. *Buildings*, 14(6). <https://doi.org/10.3390/buildings14061730>
- Zhou, B., Zhao, H., Puig, X., Xiao, T., Fidler, S., Barriuso, A., & Torralba, A. (2018). *Semantic Understanding of Scenes through the ADE20K Dataset*. <http://arxiv.org/abs/1608.05442>

Appendices

Appendix A

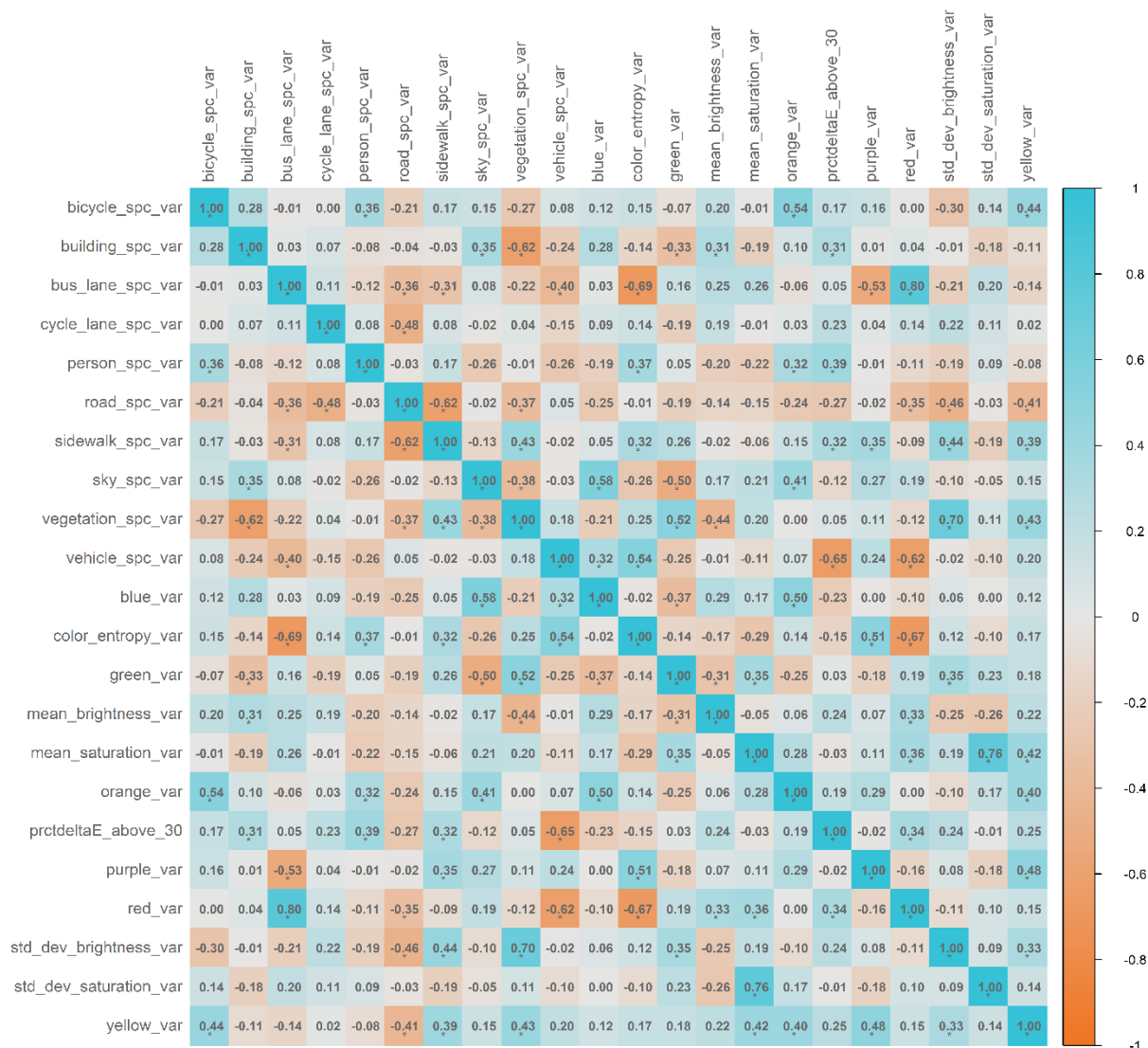
Bivariate analysis: perceptions x colour and segmentation variables (Spearman's rho)



* Significant correlation (p-value < 0.05).

Appendix B

Bivariate analysis: correlation between all colour and segmentation variables. (Pearson's r)



* Significant correlation (p-value < 0.05).

Appendix C

SEM effects directional ranking per latent construct (Functionality, colour system).

