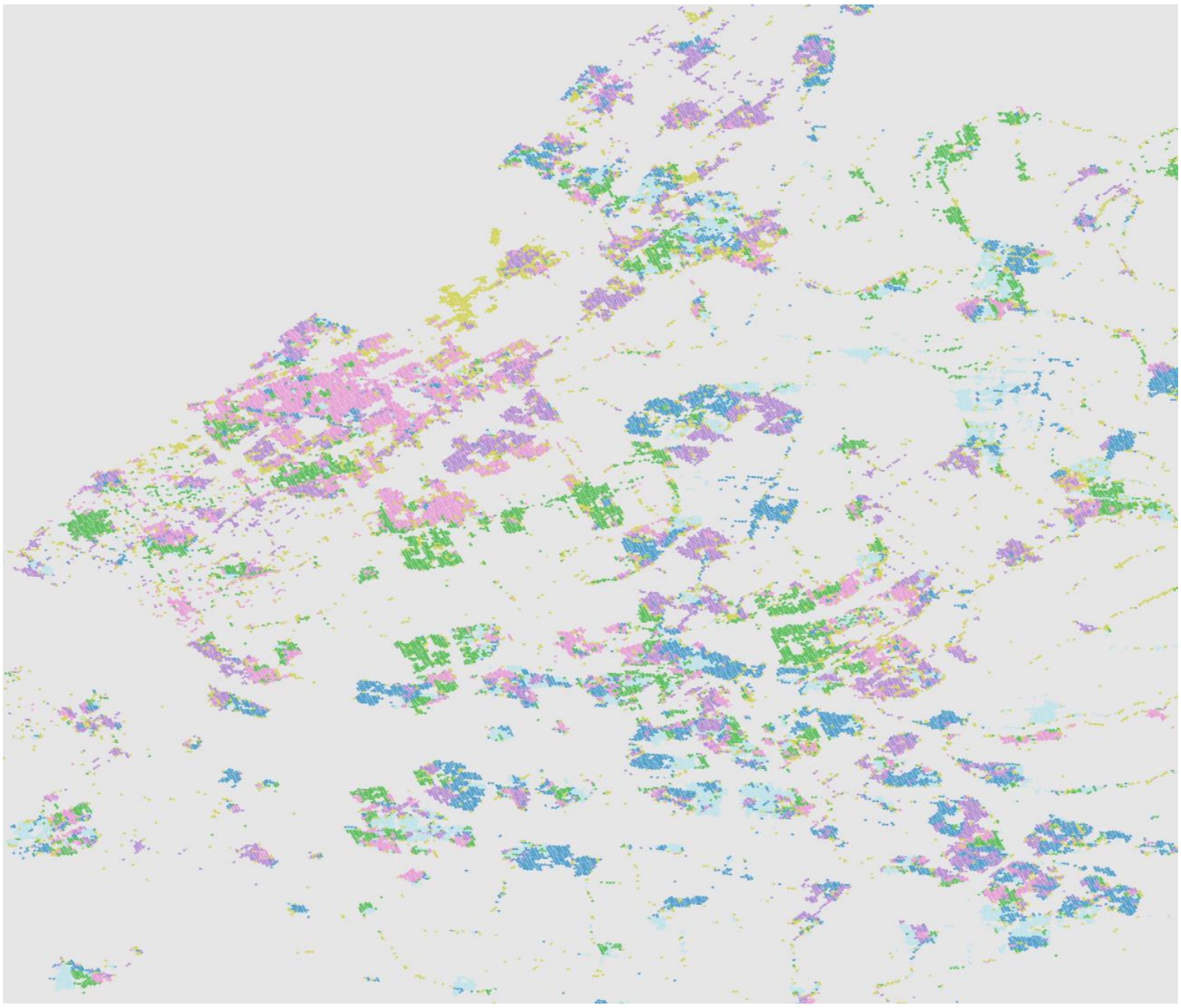
MSc thesis for Transport Infrastructure and Logistics

Learning urban representations to operationalise liveability

A transport policy perspective

Bert Berkers

2024



Learning urban representations to operationalise liveability

A transport policy perspective

Master Thesis

Submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of

Master of Science

in

Transport Infrastructure and Logistics

by

Bert Berkers

To be publicly defended on September 17th 2024

Composition of the graduation committee:

Dr.Ir. S van Cranenburgh	TU Delft, TPM, *chairperson – 1st supervisor*
Dr.Ir. A.J. Pel	TU Delft, CiTG, *2nd supervisor*
F.O Garrido Valenzuela	TU Delft, TPM, *supervisor*
Dr. J.A. Annema	TU Delft, TPM, *additional examiner*

Glossary

Term	Definition	Relevance to Transport Policy
Active Inference	A framework in neuroscience and cognitive science that describes perception, learning, and decision-making as a process of minimizing surprise and uncertainty.	Provides a dynamic perspective on how individuals perceive and interact with transport systems, shaping their travel behaviour and preferences.
Affordance	The potential for action that an environment or object provides.	The choice set from which actions are selected.
Agglomerative Clustering	A hierarchical clustering method that starts with individual data points and progressively merges them into clusters.	Useful for analyzing and grouping similar spatial units like H3 hexagons.
Bayesian Inference	A method of statistical inference that updates the probability of a prior hypothesis as more evidence becomes available.	Used in active inference models to describe how individuals update their understanding of the world by purposeful sampling through action.
Broad Prosperity	A policy evaluation framework that goes beyond traditional economic indicators to include social and environmental well-being.	Encourages a more holistic approach to transport policy, considering its impact on various aspects of quality of life.
Characteristic State	The attractor states in the reciprocal relationship between the resident and the environment. Analogous to needs/desires and preferences.	Individual preferences bias actions towards certain lifestyles: a frequent public transport traveller is more likely to live closer to a train station than a frequent car driver.
Circle Loss	A type of loss function used in machine learning, particularly for metric learning tasks.	It can be applied in urban representation learning to improve the quality of learned representations compared to triplet loss.
Cognitive Security	The protection of mental processes and the ability to construct one's niche independently from external manipulation or interference.	Relevant in developing niche constructing living digital twins.
Constrained Maximum Entropy Principle	A principle in statistical mechanics and information theory that states a system will maximize its entropy while satisfying given constraints.	Applies to location-based accessibility, considering both the attractiveness of locations and the constraints of the transport system.
Convolution Neural Network (CNN)	A type of deep learning network commonly used for analyzing visual imagery.	Used in processing street-view or satellite images for urban analysis and transport planning. See also spatial convolutions in our novel Ring Aggregation methodology.
Cost-Benefit Analysis (CBA)	An economic evaluation method that compares the costs and benefits of a project or policy in monetary terms.	Widely used in transport planning to assess the economic viability of infrastructure projects and policies,

	Relying on valuation as seen in the static approach to liveability.	helping decision-makers determine if the benefits outweigh the costs.
DGGS (Discrete Global Grid System)	A spatial reference system that uses hierarchical tessellation of cells to represent the Earth's surface.	Provides a framework for representing and analyzing spatial data in transport planning, enabling efficient computation and visualization.
Dynamic Approach (to Liveability)	An approach that views liveability as an ongoing, adaptive process between residents and their environment.	Encourages transport planners to consider the evolving relationship between transport systems and urban liveability.
Ecological Psychology	A psychological approach that emphasizes the study of behavior in the context of the environment.	Offers a perspective on how individuals perceive and interact with their transport environment, influencing their travel choices and behavior. Rather than objectively evaluate the choice set at every decision moment, one wishes for external events to occur proactively.
Enactivism	A theory in cognitive science that views cognition as a dynamic interaction between an acting organism and its environment.	Challenges traditional views of transport behaviour by highlighting the role of embodied experience and interaction with the environment in shaping travel choices.
Expected free energy	Information-theoretic quantity, which expresses the potential of an action to cause expected/unsurprising future observations.	It can be decomposed into pragmatic and epistemic values. Pragmatic value is better known as utility, and epistemic value is like option value.
Free Energy Principle	A physical principle which postulates that biological systems minimize their free energy to maintain their order and survive.	Provides a theoretical foundation for understanding the dynamic relationship between individuals and their environment, explaining how they adapt and make decisions to minimize uncertainty.
Generative Model	A type of statistical model that can generate new data instances through sampling.	In transport, these models can simulate and predict travel behaviour, enabling scenario planning and policy evaluation.
H3 Geospatial Index	Uber developed a hierarchical hexagonal geospatial indexing system.	Useful for efficient spatial analysis and representation of transport-related data at various (nested) scales.
Hedonic Pricing	A method for estimating the value of environmental amenities or disamenities based on their impact on property prices.	It can be used to assess the economic value of transport infrastructure and its impact on liveability and quality of life.
Homo Economicus vs. Homo Narrans	Contrasting views of human behaviour, where Homo Economicus is a rational, utility-maximizing agent, while Homo Narrans constructs meaning through narratives and experiences.	Highlights the limitations of traditional economic models in transport planning and emphasizes the importance of considering the subjective and narrative aspects of travel behaviour.

Indicator	A measurable variable which is used to represent or quantify a concept or phenomenon.	In transport policy, indicators are used to assess and monitor various aspects of the transport system, such as accessibility, safety, or liveability.
Information Gain	The reduction in uncertainty or entropy is achieved by acquiring new information.	In active inference, it guides the selection of actions that are expected to resolve the most uncertainty. See also option value.
Leefbaarometer	The Dutch government commissioned an instrument to measure and monitor liveability across the country.	A practical example of how liveability is operationalised in transport policy evaluation in the Netherlands.
Living Digital Twin	A dynamic, data-driven representation of a physical system that can simulate and predict its behaviour.	It can be used to model and optimize transport systems in real time, improving decision-making and policy implementation.
Location-based Accessibility	A measure of the ease with which activities or opportunities can be reached from a given location.	Central to transport planning, helping evaluate the effectiveness of transport systems in connecting people to destinations.
Markov Blanket	In probability theory, the set of variables that shield a variable from the influence of other variables in a Bayesian network.	In active inference, it separates internal and external states, enabling the system to maintain its integrity and interact with the transport environment.
Mental Representation	Internal symbols or connections that correspond to external reality.	Influences how individuals perceive and interact, affecting their travel decisions.
Multi-modal Learning	Machine learning techniques that can process and relate information from multiple types of input or "modalities".	Enables the integration of diverse data sources (e.g., images, text, sensor data) in transport analysis and planning.
Multi-Criteria Analysis	A decision-making tool that evaluates multiple, often conflicting, criteria to support decision-making.	It can be used in transport policy to weigh different objectives, such as accessibility, liveability, and safety when evaluating infrastructure projects.
Needs/Desires	Drivers of behaviour. Analogous to characteristic states and preferences.	In transport policy, fulfilling needs and desires is utility maximisation.
Niche	The reciprocal relationship between resident and environment, in its totality.	The personalised choice set with which travellers go out into the world. Perceiving only those affordances with which they are familiar and habituated.
Niche Construction	The process by which organisms modify their environment to meet their characteristics states.	In transport, it highlights how individuals and communities shape their transport environment through their choices and behaviours. See residential self-selection.
Operationalisation	The process of defining abstract concepts in terms of observable, measurable variables.	Essential for translating theoretical concepts of liveability and accessibility into practical metrics for transport policy evaluation.

Option Value	There is value in having multiple options regardless of the content of those options.	In transport networks, having multiple similar destinations and redundant network links improves the chance that the needs/desires of a traveller can be satisfied.
Percept	A latent representation of an external thing in the world.	The product of perception. Operationalised using neural networks.
Predictive Processing	A theory in cognitive science suggests that the brain constantly generates and updates predictions about sensory input.	It provides a framework for understanding how individuals anticipate and respond to changes in their transport environment, which influence their travel behaviour and decision-making.
Renormalising Generative Model	A type of generative model that can handle multiple scales or levels of abstraction.	Useful for modelling complex transport systems that involve interactions across different spatial and temporal scales. Aligns with the H3 DGGS.
Representation Learning	A machine learning approach that aims to learn meaningful and useful representations of data, often in a lower-dimensional space.	Used in urban representation learning to create compact and informative representations of urban environments from high-dimensional data, supporting transport planning.
Ring Aggregation	A method for aggregating spatial data based on concentric rings around a central point.	It can be applied to analyze the impact of transport infrastructure on surrounding areas at various distances.
Self-supervised Learning	A machine learning paradigm. The model learns the inherent structure of the data without specifying the desired target but with discrete labels, like spatial units in a DGGS.	Enables the learning of useful representations from unlabeled urban data, reducing the need for manual annotation and facilitating data-driven approaches to transport planning.
Spatial Convolutions	A type of mathematical operation used in image processing and deep learning to extract features from local neighbourhoods in images or spatial data.	Applied in urban representation learning to capture spatial relationships and patterns in urban data, enabling the learning of context-aware representations for transport analysis.
Static Approach (to Liveability)	An approach that views liveability as a measurable outcome at a specific point in time.	Commonly used in current transport policy evaluation but may miss the dynamic aspects of urban systems.
Triplet Network	A type of neural network architecture used in metric learning, which learns from triplets of examples. Minimising the difference between anchor and positive, maximising the difference between anchor and negative.	It can be used to learn similarities between different spatial units.
Urban Representation Learning	The application of machine learning techniques to create compact, meaningful representations of urban environments.	Enables the automation of liveability assessment and supports data-driven approaches to transport planning and policy evaluation.

Utilitarianism	A philosophical approach that evaluates actions based on their overall utility or happiness for the greatest number of people.	Often used in transport policy evaluation, but its limitations in addressing equity and distributional concerns have led to the exploration of alternative frameworks like broad prosperity.
Valuation	The process of assigning a value or worth to something. It is often used in economics and policy evaluation.	In the context of liveability, it refers to how residents perceive and value different aspects of their living environment, including transport infrastructure and services. The more needs are satisfied, the better.
Variational Free Energy	A quantity in the Free Energy Principle that measures the difference between an organism's internal model and its sensory inputs.	In transport modelling, it can represent the discrepancy between an individual's expectations and actual experiences of the transport system, driving adaptive behaviour.

Contents

Preface

This thesis is the result of the Master Transport Infrastructure and Logistics, a period in which I learned not only about transportation but about infrastructure in its broadest sense—more than static physical artefacts but animated by a dynamic process. Liveability is ultimately the study of such an animating force, whether that of residents, cities, or countries.

Making digital twins with hexagons allows me to combine two games played throughout life, Maxis's SimCity and Sid Meier's Civilization series. At some point, simulation is insufficient to describe the real world. Ironically, this has led to the study of even more simulations.

Several courses in the master's program (policy track) provided the foundation to consider the dynamic approach. Most notably, the repeated focus in many courses on the limitations of utilitarianism led me to search for alternatives. The active inference institute provided ample affordances to pursue one such alternative.

Today's rapid technological progress means that this thesis is an attempt to keep up. Some terminology, like simulating systems' niche construction process, may seem fantastical, but it is already being implemented in other contexts, such as business, regenerative agriculture and the spatial web. At the same time, there are seemingly a dozen crises, each relating to transportation in some way, shape or form.

I am grateful to my family, friends, and flatmates for their supportive environment and the mentorship offered by the supervisors. The level of writing achieved in this thesis would not have been possible without the constructive patience of all involved.

Throughout the writing of this thesis, I used the language applications Grammarly and Claude 3.5 Sonnet. For all text to which it applies, Claude was first extensively prompted by me using drafts of the thesis and then further edited. No new ideas originated from Grammarly or Claude. They are simply not good enough yet except for restructuring messy drafts, providing feedback, and assisting with Python code.

Summary

Cities worldwide face the dual challenge of increasing density while developing sustainably. These pressures strain transportation systems, necessitating optimal performance. In the Netherlands, transport policy focuses on three main objectives: accessibility, liveability, and safety. This thesis addresses accessibility and liveability.

Liveability concerns the fit between residents and their living environment. This fit is an ecological relationship involving interactions with the built, natural, and social environments. The thesis explores two approaches to operationalising this ecological fit: static and dynamic. The static approach, exemplified by tools like the Leefbaarometer, views liveability as a measurable outcome. It considers the valuation of the environment as a proxy for fit, obtained through surveys or hedonic pricing methods. This approach is widely used in current transport policy evaluation but is limited in its ability to do ex-ante analysis.

On the other hand, the dynamic approach views liveability as an ongoing process. It considers fit the difference between what a resident expects to observe and experiences. This approach uses generative modelling, which involves causal probabilistic relationships. The manner in which these generative models are updated is unique, attempting to capture the dynamics of the reciprocal relationship between the resident and the living environment. The resident embodies the generative model, containing both internal and external states. The generative model is definitive of a resident's niche. Construction of the niche is afforded by accessibility.

Objective

The thesis aims to work towards future operationalisation of the dynamic approach to liveability. While full implementation is far beyond the scope, the objective is to understand the role of representations in defining and operationalising liveability. In the context of transport systems, representations can be understood as simplified and compressed digital twins of complicated urban environments. These representations can be used for decision-making in infrastructure planning and policy formulation.

The dynamic approach to liveability relies on the action-perception loop as its generative model. This loop means that perception is anticipated, an action in itself. Selecting optimal actions is a matter of maximising epistemic and pragmatic value. Writing a thesis is an action; hence, we follow the formal decomposition of action selection:

- Epistemic component: develop a theoretical framework, exploring the action-perception loop in liveability and the role of representations within this loop. Bridging cognitive science and transport modelling enables novel machine learning applications.

- Pragmatic component: automate the operationalisation of liveability by applying urban representation learning techniques to the province of South Holland. Our novel contribution is the developed mathematical model based on spatial convolutions. The study uses various data sources relevant to transport planning, including road networks, public transport schedules, aerial and street view images, and points of interest.

Modelling Study

Addressing the pragmatic component, we develop a novel urban representation learning technique using the H3 hierarchical hexagonal discrete global grid system (DGGS). Neural networks are used to extract representations such that each discrete hexagon is assigned a coordinate in the metric representation space. Features from different data sources are extracted using individual encoder networks. These features are then combined using a late-fusion network to obtain the final representations. We develop Ring Aggregation, a mathematical model to fuse multiple features while accounting for spatial context through sampling heuristics and spatial convolutions. The methodology considers insights from the dynamic approach to liveability, like landscapes of affordances, analogous to location-based accessibility. No operationalisation of the dynamic approach is developed in this study. Instead, this study applies the static approach and evaluates urban representations using multiple univariate linear regressions with Leefbaarometer scores as targets.

Key Epistemic Observations

- The action-perception loop is central to the dynamic approach to liveability but absent in the static approach.
- Representations play an instrumental role in the action-perception loop, informing action selection.

Key Pragmatic Observations

- Regarding sampling heuristics used to train the late-fusion network, Euclidean distance and location-based accessibility perform similarly, better preserving urban area integrity.
- The configuration of Ring Aggregation significantly impacts urban representation quality, showing heterogeneity between Leefbaarometer scores. Socially oriented scores can perform just as well with smaller receptive fields and steeper weighted average functions like exponential.
- Different data sources excel in predicting various aspects of liveability, suggesting the importance of integrated data approaches in transport policy.
- Compared to urban2vec and M3G studies, Ring Aggregation outperforms them in predicting Leefbaarometer scores across the board.

Looking Forward

By reframing the roles of indicators, perception and preferences, this thesis concludes that static and dynamic approaches to liveability are complementary. Complementarity implies that the static approach to liveability can be used to bootstrap the development of the dynamic approach. While 'living' digital twins operationalising the dynamic approach do not yet exist, this study provides building blocks towards their development. Future directions for research and application in transport policy centre on the hypothesis that hierarchical active inference models are already transport models, just waiting to be used.

1 Introduction

This thesis addresses the definition and operationalisation of liveability in transportation policy. It finds an underdeveloped approach to liveability and addresses it through research into literature from cognitive science and conducting a modelling study. The latter of which aims to provide technical prerequisites for this underdeveloped approach. Finally, directions for future research are made to contextualise liveability into broad prosperity as studied in transport policy.

1.1 Societal Relevance

Many cities worldwide face challenges in urban development, balancing the need for housing with sustainable development (United Nations Department of Economic and Social Affairs, 2023). Two primary pressures on the transport system have emerged. First, densification strategies adopted to address the housing crisis offer benefits for sustainable urban living but also increase pressure on existing transportation systems and public spaces. Second, there is a push for sustainable modes of travel, aligning with emission reduction targets for 2050 (Ministerie van Inrastructuur en Waterstaat, 2023). Dense urban environments promote environmentally friendly travel modes such as walking, cycling, and public transport, creating a synergy between densification and sustainability goals (Gupta et al., 2024).

Concurrent with these two pressures, transport policy has three significant shifts. 1) A systems dynamics approach now recognises cities as complex, self-organising systems. This understanding has led to considering multiple leverage points beyond traditional infrastructure investments, including land use planning, demographic shift management, and behavioural nudging (Huibregtse, 2021; Ministerie van Inrastructuur en Waterstaat, 2023; Reudink et al., 2023). 2) Simultaneously, interdisciplinary planning has become crucial due to the scarcity of public space in dense urban environments. Limited space must accommodate climate adaptation measures, areas for social interaction, infrastructure for healthy travel options, green spaces for well-being, and solutions for reducing pollution (Snellen & Bastiaanssen, 2021; Ståhle, 2008). The ongoing energy transition, particularly vehicle electrification, adds further complexity to these urban planning challenges. 3) Finally, transport policy planning is shifting from a narrow economic focus to a broader understanding of societal well-being (Raad voor de leefomgeving en infrastructuur, 2024; Snellen & Bastiaanssen, 2021). Broad prosperity expands the scope of policy evaluation beyond monetary valuation, additionally focusing on what people are capable of and an equitable distribution of benefits and costs.

1.2 Liveability in Transport Policy

Liveability has become a key consideration in Dutch transport policy and is now considered alongside traditional objectives such as accessibility and safety (Huibregtse, 2021). Liveability is generally defined as the fit between residents and their living environment (Dorst, 2005).

However, operationalising liveability for policy evaluation presents significant challenges. Current approaches to measuring and operationalising liveability often rely on static interpretations, viewing it as a measurable outcome rather than a dynamic process. Tools like the Leefbaarometer in the Netherlands exemplify the static approach (Mandemakers et al., 2021).

While practical as a measurement tool for policy-making, these static measures may not fully capture the complex, reciprocal relationship between residents and their environment. For example, since the static approach does not consider causal relationships, it cannot be used for ex-ante assessment of interventions, only to be used as a signalling instrument (Mandemakers et al., 2021). Furthermore, static models are often estimated for large study areas to improve statistical confidence in parameter values. However, this also means that a subjective concept like liveability is generalised for an entire population, e.g. all Dutch people.

Leidelmeijer (2004) recommended further study into liveability as a longitudinal dynamic process, emphasising an ecological perspective. This ecological approach views liveability as an ongoing interaction between residents and their environment, where both continuously influence and shape each other. It also aligns with behavioural geography, which studies how individuals perceive, interact with, and adapt to their surroundings over time (Acker & Witlox, 2008; Barker, 1978). Besides liveability, sustainability also involves an ecological relationship with the world (Dorst, 2005). Liveability concerns local spatio-temporal scales, whereas sustainability covers the globe, spanning many decades into the future.

The dynamic approach to liveability, which defines the fit between residents and their environment as an ongoing, adaptive process, resonates with transport policy. Central to this dynamic perspective, the action-perception loop reflects the continuous interaction between residents and their urban surroundings. It posits that individuals react to their environment and actively shape it through their choices and behaviours. The concept of niche construction, where residents actively modify their surroundings to create suitable living spaces, further underscores this relationship's dynamic and reciprocal nature.

In the context of transport policy, this translates to a shift away from viewing transport systems as mere facilitators of movement towards recognizing them as integral components in creating liveable urban environments. The accessibility afforded by transport networks, for instance, shapes the 'landscape of affordances' that residents navigate, influencing their choices and behaviours. The dynamic approach also sheds light on phenomena like residential self-selection, where individuals gravitate towards neighbourhoods that align with their travel preferences, and the flow of ideas within communities facilitated by transport networks that enable social interaction and innovation. It encourages a move beyond isolated infrastructure projects towards a more holistic understanding of how transport systems can be designed to foster and support the dynamic processes that contribute to a high quality of urban life.

1.3 Research Objective

The limitations of current static approaches to liveability highlight a significant research gap. There is a need for a more dynamic, process-oriented understanding of liveability that can account for the continuous interaction between residents and their urban environment. This dynamic approach aligns with ecological perspectives in behavioural geography (Aitken & Bjorklund, 1988) and recent developments in cognitive science, particularly concerning multi-scale action-perception loops (M. J. D. Ramstead et al., 2019), which describes the reciprocal relationship between actors and their environment, see Figure 1. When actors share the same option set of actions, they share a niche.

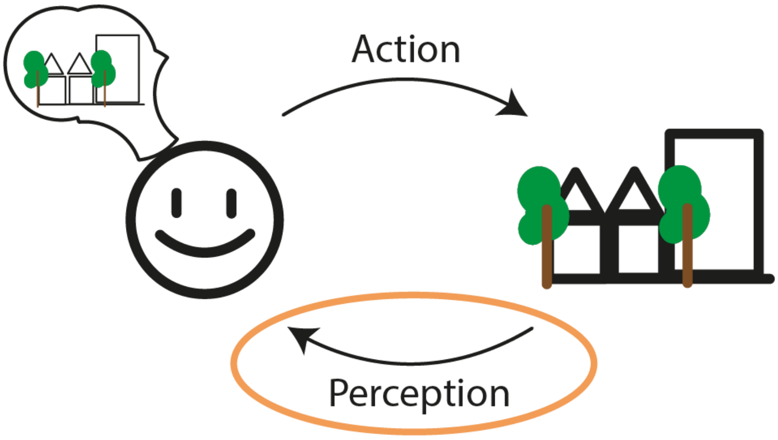


Figure 1: The action-perception loop describes the reciprocal relation between a resident and its living environment. The resident acts on the environment by moving within or changing it, and vice versa, perceives the environment, creating a (mental) representation thereof. The representation is optimistically biased towards one's preferences (K. Friston et al., 2013). Liveability minimises the difference (maximising fit) between this mental representation and the environment, mediated by action and perception.

In line with the decomposition of action selection in the dynamic approach, we delineate the research objective into two components: pragmatic- and epistemic value. Epistemic value is taken care of by developing the dynamic approach and contrasting it to the currently dominant static one, addressing the gap identified by Leidelmeijer (2004). Pragmatic value, on the other hand, relates to modelling considerations. The static approach relies on indicators and valuation (surveys), which require resources to collect and process. The dynamic approach to liveability addresses the need to gather this data by automating its processing and discarding the need for valuation altogether. Instead, the dynamic approach attains fit by introducing information engines, which produce work by achieving and maintaining their niche through (mental) action. The niche is biased towards its characteristic states, better known as preferences, see Figure 1. This thesis aims to study the development of the transmission for such an information engine by drawing upon representation learning. The transmission is the mapping between high-dimensional data in the urban environment and the compressed computer-readable representation of that data, which can then be used in a subsequent information engine. Of course, the obtained representations can also be used in the static approach; however, one would make different modelling decisions for optimal performance. So, the method is to learn representations tailored to the dynamic approach while verifying them using a static operationalisation.

Representations are essential in the action-perception loop as they are used to reason about the world and select optimal actions. For transport policy aims, the object to be represented is the urban region. Hence, we draw upon the urban representation learning literature. Urban representation learning may be used for static and dynamic approaches to liveability, offering a versatile tool for analysis. On the other hand, active inference provides a framework specifically suited for multi-scale action-perception loops—providing the tools to create models that themselves construct models of the world (Ramstead et al., 2024). In consideration of both epistemic and pragmatic value, the research objective is as follows:

To understand the role of representations in defining and operationalising liveability.

To address the objective, we construct a theoretical framework to define and contrast two approaches to liveability. Additionally, we conduct a modelling study to address our limited understanding of operationalising the dynamic approach to liveability, focussing on learning representations ‘as if’ they would be used as transmission in an information engine. To this end, the urban representation learning literature will be outlined, and a novel learning strategy will be developed considering the dynamic approach to liveability.

We identify six research questions. The first two relate to defining liveability, drawing extensively on the action-perception loop. This loop is the basis of the ecological relationship between residents and their living environment. The ecological relationship makes it possible for a fit between them to exist in the first place. The role of representations is subsequently aimed at contextualising the operationalisation of fit. The latter four research questions relate to the operationalisation of representations. Note that there will be no operationalisation of fit, as that requires the information engine. Instead, this thesis tackles the transmission, such that urban representation learning is used to map the input data towards a compact representation. The latter may be used as input for future work that develops the total assembly comprised of transmission and engine (Figure 24). How these urban representations are learned depends on several modelling choices: how to account for proximity, local context, and data sources. Where proximity and local context attempt to capture the first law of geography, which states that similar things are found closer togheter (Tobler, 1970). Proximity tackles this by addressing the sampling heuristic used to train the neural networks. While accounting for spatial context involves the architecture of the neural network by introducing spatial convolutions. Finally, a comparison to two other urban representation learning studies is conducted (Huang et al., 2021; Z. Wang et al., 2020).

Identifier	Question	Methodology
RQ1	What is the role of the action-perception loop in liveability?	Literature Review
RQ2	What is the role of representations in the action-perception loop?	Literature Review
RQ3	What is the impact of the chosen proximity measure in the sampling heuristic used to calculate similarity loss?	Modelling Study
RQ4	What is the impact of configuration on aggregating over the local spatial context?	Modelling Study
RQ5	What is the added value of different data sources?	Modelling Study
RQ6	What is the impact of learning strategy?	Modelling Study

1.4 Thesis Structure

This thesis examines the role of representations in defining and operationalising liveability, progressing from theoretical foundations to empirical applications of urban representation learning.

Chapter 2 establishes the theoretical framework, delineating static and dynamic approaches to liveability. It introduces key concepts, including the action-perception loop and niche construction, contextualizing these within transport phenomena such as residential self-selection and idea flow.

Chapter 3 outlines the methodological approach, detailing data collection processes, model specifications, and experimental design. The chapter focuses on the novel ring aggregation method, which utilizes the H3 hexagonal DDGS and spatial convolutions to capture spatial relationships in urban environments.

Chapter 4 presents the results of urban representation learning experiments, addressing each research question systematically. It encompasses visualisations of agglomerative clustering and graphs showing the predictive accuracy of Leefbaarometer scores.

Chapter 5 synthesizes the key findings, concisely answering the research questions and evaluating the main objective. It assesses the potential of urban representation learning to automate and enhance liveability assessment in transport policy contexts.

Chapter 6 discusses broader implications, examining the potential of hierarchical active inference models in transport modelling. It notes the similarity between the H3 hierarchical DGGS and the newest generative models. Drawing parallels between message passing in these generative models and travel journeys in transport networks. The chapter concludes by outlining future research directions, including the development of 'living' digital twins for urban environments and addressing the mesoscale sustainability gap. To bridge the ecological relationship, from liveability at local spatio-temporal scales towards sustainability at global scales.

The thesis includes three appendices: a condensed version of the research in the format of a scientific paper, technical details of the learning strategies, and a dictionary of point-of-interest labels used in the study.

2 Theoretical framework

The theoretical framework explores urban liveability, its definition, and operationalisation. It draws from urban planning, transportation policy, ecological psychology, and cognitive science to provide a comprehensive view of liveability. The framework addresses the ecological relationship between residents and their urban environments.

The framework delineates two distinct approaches to understanding liveability. The static approach is based on utilitarianism and conceptualises fit as a matter of valuation and satisfaction. It views liveability as a measurable outcome that can be quantified through indicators and surveys. In contrast, the dynamic approach is grounded in surprise minimisation and views fit as a process of niche construction and achieving optimal grip on the world. This approach considers liveability an ongoing, adaptive process between residents and their environment.

By distinguishing and contrasting these two perspectives, the framework provides a comprehensive understanding of urban liveability, encompassing traditional measurement-based methods and emerging dynamic, process-oriented approaches. Furthermore, aligning with differences in world views: *'homo economicus'* versus *'homo narrans'*, where people are either rational choice makers or narrative constructors of their lived embodied experience.

Static liveability is conceptualised through a perspectival measurement of the reciprocal relationship between resident and environment. Focusing on the resident's perspective of the environment is called liveability, while the reverse is quality of life. A second axis is that of the spatiotemporal scale considered. Liveability is local and now, whereas sustainability is global and into the future. Both axes of this perspectival measurement have intermediates. Environmental quality is a more objective term outside these two perspectives, as it does not include subjective perception. No perspective is taken.

Dynamic liveability emphasises perception as a bidirectional process. The approach is rooted in ecological psychology and introduces concepts like niche construction and affordances. The niche is viewed as the generative model of the coupled reciprocal system. Unlike the static approach, which separates residents and the environment, the dynamic approach considers the complete system as one entity of interest.

The operationalisation of static liveability encompasses three key approaches: perceived, indicated, and apparent liveability. These methods typically involve fixed indicators and metrics, such as those used in the Leefbaarometer. Recent advancements in urban representation learning offer promising avenues to automate this process while also approximating human perception, potentially providing more expressive alternatives to traditional labour-intensive indicators.

While the operationalisation of dynamic liveability is still theoretical, active inference models show potential for modelling the fit between residents and the environment. These models can themselves model, aligning with the dynamic nature of liveability. Representations used in static operationalisations can be incorporated into dynamic models, but their role shifts from being outcomes to study to being instrumental in the process. Dynamic operationalisation focuses on the behaviours emerging from these representations rather than the representations themselves.

Urban representation learning, the study of mapping high-dimensional urban data such as street view images onto lower-dimensional representations, applies to both approaches. Both approaches rely on indicators, percepts thereof, and needs/desires. Instead of conventional indicators, one can use high-dimensional data, feeding it into a neural network that extracts representations/percepts.

Table 1: Readers' guide to the theoretical framework.

	Static	**Dynamic**
Definition	Perspectival measurement (0 Defining Liveability) Valuation (0 Defining Liveability) Environmental Quality (0 Defining Liveability) Quality of Life (0 Defining Liveability) Homo Economicus (2.2 Liveability in Transport Policy Evaluation)	Action-perception loop (2.4 Ecological Liveability) Free energy minimisation (2.4.2 Formalising the Dynamic Approach) Niche construction (2.4 Ecological liveability) Affordances (2.4 Ecological liveability) Homo Narrans (2.2 Liveability in Transport Policy Evaluation)
Operationalisation	Perceived liveability (2.3 Operationalising Urban Liveability) Indicated liveability (2.3 Operationalising Urban Liveability) Apparent liveability (2.3 Operationalising Urban Liveability) Leefbaarometer (2.3 Operationalising Urban Liveability) Urban representation learning (2.8 Urban Representation Learning)	Active inference models (2.4.2 Formalising the Dynamic Approach) Instrumental role of representations (2.5 Mental Representations) Perception as action (2.5.3 The Role of Representations in Action-Perception Loops) Urban representation learning (2.8 Urban Representation Learning)

2.1 Defining Liveability

Liveability is a complex and multi-dimensional concept encompassing various aspects of urban life. It is often used interchangeably with terms like quality of life, well-being, environmental quality, and health, leading to ambiguity in its definition. The primary difficulty in defining urban liveability lies in the entangled nature of these concepts, each studied by distinct disciplines. For instance, environmental scientists simulate the propagation of sound and pollution across urban regions, while health researchers examine outcomes from lifestyle choices such as active mode use or the presence of healthy eating options. Sociologists study general life satisfaction to understand individual components' contribution to one's life, often termed quality of life. Lastly, urban planners focus on the living environment and how it fits the needs and desires of those residing there. Quality of life and liveability are commonly used interchangeably (Tan et al., 2024). Both these terms aim to capture the extent to which an environment meets the needs of its residents. As De Haan et al. (2014) put it, "*Fulfilling human needs, and fulfilling more of them, increases the quality of life.*"

Leidelmeijer (2004) provides a succinct demarcation of terminology, starting with the foundational definition that liveability is the fit between a resident and their living environment. This fit should be understood as occurring within a dynamic, coupled process between the environment and residents, meaning that urban liveability should be seen as a process rather than an outcome. In line with (Veenhoven, 2000), this fit is understood through an ecological approach, with Veenhoven stating, *"liveability is the fit of the environment on the adaptive capabilities of a lifeform"*. Measuring this entangled nature in practice is challenging because all variables are intertwined in this dynamic reciprocal process; the environment impacts the resident, and vice versa. Figure 2 illustrates this concept using a Venn diagram highlighting the entangled process. By focusing on specific components within this conceptual model, definitions currently operationalised are derived: when focusing on the environment, it is termed liveability; when focusing on the residents, it is termed quality of life. Sustainability and liveability are related but distinct in their scales. Both concern meeting needs (De Haan et al., 2014). However, where sustainability concerns a much larger spatiotemporal vista than liveability, the latter is reserved for local neighbourhood-scale relations between residents and their daily local environment—the here and now. Sustainability may resemble the ecological relationship between cities and the world over the years or even decades, concerning the needs of collectives rather than individuals.

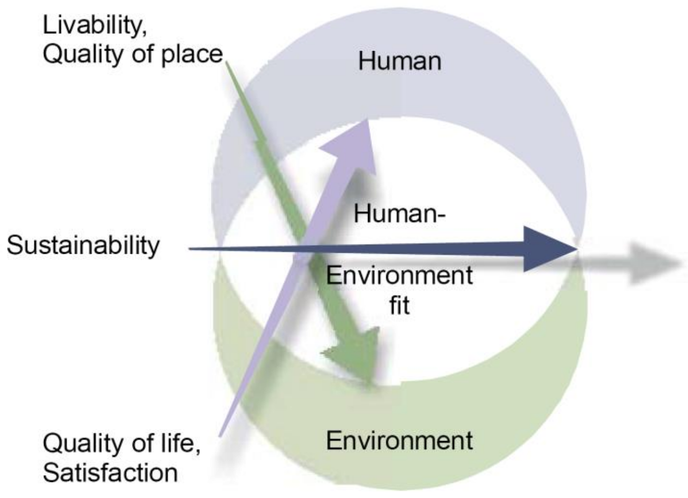


Figure 2: Perspectives on human-environment fit viewed as a Venn diagram. From (Leidelmeijer, 2004).

Operationalising the reciprocal relationship between residents and the environment statically through perspectival measurement is not the only available approach. Aitken & Bjorklund (1988) aimed to make the transactional perspective in behavioural geography accessible, which starkly contrasts the act of dividing the reciprocal relationship into measurable components for subsequent study.

The static approach views residents' inner world as static and measurable outcomes given their environment. In contrast, the dynamic approach considers the reciprocal relationship between residents and their environment as the system of interest. Snipping this reciprocal relationship in half characterises a static interpretation of urban liveability, whereas the literature emphasises the validity of the dynamic approach, illustrating the discrepancy between theory and practice.

Indeed, as (Leidelmeijer, 2004) notes, the transactional view of behavior-environment relations may be the most compelling approach to conceptualising urban liveability and deserves further study. This view is concurrent with the ecological perspective, which considers the world made of interconnected dynamic systems. The relationship between these systems and the unfolding process interests the analyst. Therefore, the definition of liveability used in Dutch literature—the fit between environment and resident—should be understood from an ecological, transactional perspective (Figure 3). However, Operationalising this definition requires measurement, meaning that the analyst must take a perspective, either focusing on the human or the environment, thus modelling quality of life or urban liveability, respectively.

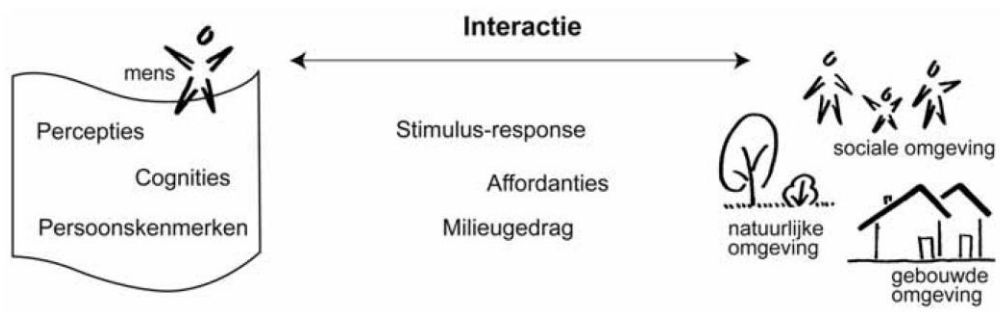


Figure 3: Fit as the product of interaction in an ecological relationship. From (Dorst, 2005).

Consequently, current operationalisations of urban liveability do not model the interaction of liveability in its ecological sense. While the fit between resident and environment is dynamic, a perspectival measurement of this process is taken. Practical considerations have led to alternative definitions. For instance, Valk & Musterd (1998) posit that *"liveability is the valuation by the individual of their living environment"*, and Dorst (2005) summarises current practice as a summation of valuations for various aspects, in line with what Pacione (2003) refers to as the simplest model of life satisfaction. Leidelmeijer (2004) outlined current operationalisations within this limited definition of urban liveability by presenting Figure 4, where liveability is represented by chosen indicators/valuations and caused by various determinants, often environmental attributes. Section 2.3 will outline how different operationalisations build off the definition in Figure 4. Indeed, current operationalisations are only made possible by defining urban liveability as a measurable outcome represented by indicators and caused by determinants. In modern practice, the independent variables are indicators of proximity to schools, and the dependent ones are valuations/satisfactions.

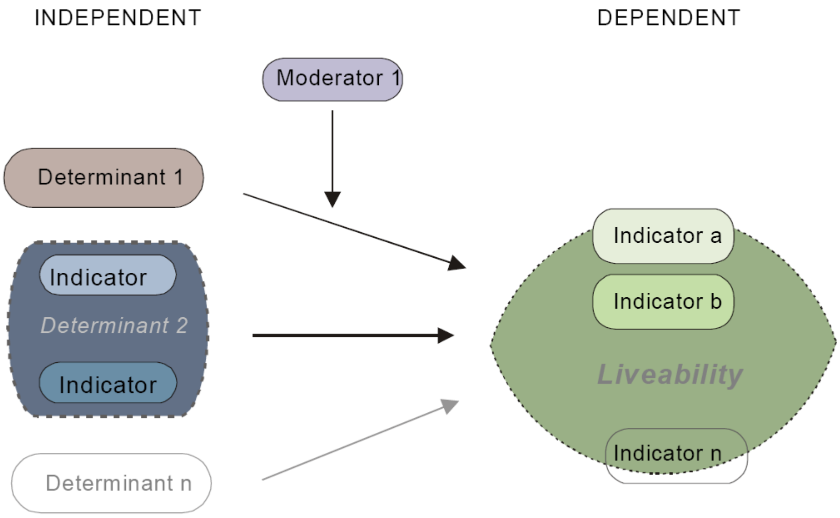


Figure 4: Operationalising urban liveability as a static outcome. From (Leidelmeijer, 2004).

The bi-directional dynamic between policymakers and researchers greatly influences the definition and operationalisations, as noted by (Valk & Musterd, 1998). Policymakers focus on actionable liveability measures, while researchers aim to capture the concept as accurately as possible. This bi-directional nature can be seen in many studies that often begin by listing various actionable aspects to be considered, thereby locking themselves into the static approach to liveability. The uniformity with which urban liveability is understood as the study of adding more variables means that the variety of effects is well-indexed and studied. Unfortunately, the validity of the impact of these effects is contestable, as current framings do not capture liveability in its ecological essence. Instead, causality in the model specification is assumed by taking a perspectival measurement of the process from the perspective of humans towards the environment—valuation.

Causality presents a difficult problem in the study of urban liveability, largely due to the dependence of operationalisations on residents' valuation of the environment. Other life circumstances, which may relate more to quality of life, can lead to a low opinion of one's living environment (Leidelmeijer, 2004; Veenhoven, 2000). Circumstances related to quality of life may be related to employment or personal relationships.

Furthermore, models of urban liveability typically assume a structurally causal model. That is, theory informs the analyst of the direction of effects, estimating interpretable and meaningful coefficients. Therefore, the Leefbaarometer (Mandemakers et al., 2021), the state-of-the-art operationalisation, explicitly notes the inherent lack of causal predictive power. It serves merely as a measure indicating what liveability may be and is intended as a signalling instrument for further studies.

2.2 Liveability in Transport Policy Evaluation

Around the transition to the 21st century, there was a concerted effort to understand urban liveability for planning purposes of the built environment. Particularly of interest was the relationship between liveability and sustainability. It is possible to design a liveable, unsustainable city and vice versa (Dorst, 2005). In transportation policy specifically, the relationship between liveability and sustainability was virtually nonexistent (Wolsink, 1998). More recently, however, sustainability has become integrated as a component of liveability within transport policy evaluation (Huibregtse, 2021). This perspective posits that liveability is just one policy outcome, along with safety and accessibility. With the shift in policy objectives towards liveability and sustainability, unified under broad prosperity, indicators have taken a central role. Multi-criteria analysis is proposed as a suitable evaluation methodology to weigh the importance of various indicators (de Abreu e Silva et al., 2023). Multi-criteria analysis is a methodology that aims to structure trade-offs transparently; however, it is limited to the scope of individual studies as weighting cannot be transferred (Annema et al., 2015).

Broad prosperity is a policy evaluation framework that aims beyond narrow prosperity, the latter constrained to the national domestic product. The Dutch government commissioned the development to address the narrow scope of economic indicators (Raad voor de leefomgeving en infrastructuur, 2024). Moving beyond narrow prosperity is aligned with the observation that developed countries find material prosperity insufficient for a high quality of life (Pacione, 1990). Narrow prosperity requires monetisation of effects, such as travel time and pollution. However, monetisation is limited in capturing the various effects relevant to modern policy evaluation, for example, being ill-suited to account for environmental concerns (Annema & Koopmans, 2015).

Additionally, experience-related aspects of transportation are hard to account for (monetise) under narrow prosperity, such as the quality of trips or places (Anciaes & Jones, 2020). Instead, broad prosperity comprises a large set of indicators, including social and environmental ones, in addition to economic performance. Performance on indicators should be satisfied equitably across the population rather than adhering to utilitarianism, which leads to aggregate improvement at the expense of increased inequality (de Boer et al., 2023). Broad prosperity as applied to transport policy is similar to the policy formulation outlined by the (Huibregtse, 2021). Broad prosperity, however, has slightly more aspects, such as splitting liveability into health and the living environment.

Liveability has become a cornerstone of modern transport policy in the Netherlands. Liveability is a boundary condition that should be satisfied, along with safety, while maximising accessibility. It is up to policymakers to weigh these three objectives (Huibregtse, 2021). Furthermore, it is important to note that indicators come in two forms: output and outcome. Output indicators relate to measurable things, like the number of vehicle kilometres driven or the percentage of modal share. Outcome indicators operationalise policy objectives, such that they measure the effects of interventions. Output indicators are meant to be used only for measurement and indicative purposes. Policy appraisal requires causal relationships; hence, outcome indicators are required. Outcome indicators require more research due to the need for an established causal relationship between measurements and effects on policy objectives.

As (Huibregtse, 2021) outlines, liveability relates to the impacts of the mobility system on the living environment via pollutants (noise/sound) or barrier formation. Accessibility refers to the extent to which residents can reach their desired destinations; it is only a means to an end. Safety is the third component of the modern policy framework, along with liveability and accessibility. It concerns not only traffic safety but also that caused by externalities such as the shipping of dangerous substances. Alternative perspectives on transport policy goals have been formulated by (Litman, 2011), outlining sustainability as an overarching goal, within which liveability is a subset. Four goals are defined: economic, social, environmental and good governance/planning. Each set of goals has several objectives and associated indicators. The sustainability goals outlined and their associated indicators may be understood as an outcome indicator as described by (Huibregtse, 2021), as there is justification for the supportive nature of these indicators towards sustainability. Per capita GDP, traffic noise levels, and smart growth development are exemplary indicators.

The difference between static and dynamic views on liveability may provide deeper insight into the reliance on indicators. Quantifying the performance of the transport system using indicators is implicitly a static approach. While this argument may be more straightforward for liveability, accessibility and safety can also be framed from a static vs dynamic perspective. Accessibility allows residents to construct their fit with the environment dynamically. It is only possible to figure out what niche one should occupy by exploring the broader habitat. That is, if it is easy to travel across the country, one might more easily figure out which parts thereof one would like to visit again or relocate to for living or work. Liveability is defined as the fit between resident and environment, ultimately a dynamic interpretation of action perception, and is directly linked to accessibility through the ability to construct such fit. Safety may be framed, in turn, as the ability to maintain that fit. Accidents significantly impede the dynamic process; one cannot attain fit if one does not exist anymore or if one's capabilities are reduced via injuries.

Agency, related to capabilities as studied in the broad prosperity framework (Snellen & Bastiaanssen, 2021) and distributive justice (Pereira et al., 2017), is essential in the dynamic approach (Aitken & Bjorklund, 1988). As discussed in the chapter on ecological liveability, the action potential may be that which is perceived in action perception loops. The action in the action-perception loop between residents and the environment is reflected in accessibility; more accurately, accessibility is the action potential. Option value, as studied in transport planning (K. T. Geurs et al., 2006), comes to mind. Choices can only be made if there are options in the first place; hence, there is value to be gained by having a variety of travel modes and routes at one's disposal.

Within the dynamic approach to liveability, there is great interest in the difference between habitual and purposeful change. In transport policy contexts, habitual change relates to everyday travel behaviour, such as driving to work. Habitual behaviour occurs in times of stability, where there is no need to change. On the other hand, purposeful behaviour becomes necessary in times of change, often taking the form of life events studied in residential choice locations (Fatmi et al., 2017). There is a broad literature on the relationship between life events and travel behaviour change, which has become a major policy lever. For example, life events are unique opportunities to promote car sharing, which can be exploited in the (re)-development of neighbourhoods (de Gruijter, 2019).

Land use plays an important role in the transport system. The movement of passengers and goods is only required due to origins, destinations and intermediate resistances. Accessibility is best viewed through the lens of activities distributed across space and time (K. T. Geurs & van Wee, 2004). Alternative lenses are utility, person based and location-based. Without getting into the details of different forms of accessibility, all have in common the proximity of destinations to origins in some way, shape or form. In the broad prosperity framework, accessibility encompasses capabilities beyond hard infrastructure. There may be a highway, but if one does not own a car, then the road is not of much use. Capabilities can, however, be extended far beyond obvious examples and account for those with disabilities or difficulty using digital services (Durand, 2019). Previous work has studied the role of affordances (capabilities) in pedestrian route finding (Vandenbroucke et al., 2013).

Lastly, circling back to broad prosperity with the active, dynamic perspective. A limitation of narrow prosperity and its associated operationalisation using cost-benefit analysis is the static nature of the analysis. The value of attributes in the environment is measured and valid as long as these stay relatively constant (Banister & Hickman, 2013). However, this is not the case during times of deep uncertainty, and cost-benefit analysis may lose its predictive power. Scenario-based planning has been proposed as an alternative (Banister, 2008). This methodology aligns with the dynamic view of action perception since one projects potential futures to take action rather than forecast them, leaving room to change policies as the world unfolds. Schwanen (2020) has approached the shift from static to dynamic approaches from a different angle, focusing on the transition from *homo economicus* to *homo narrans* and the continued re-enactment of the former in transport research. Later, it will become clear how this transition may be facilitated by moving from static to dynamic approaches to action perception and liveability. The dynamic approach to liveability, operationalised using active inference, perfectly describes *homo narrans* (Bouizegarene et al., 2024). An intuitive distinction between the static and dynamic conception of action perception is the role of preferences, where as static approaches assume full optimism towards achieving ones perferences (K. Friston et al., 2013), dynamic approaches to action perception construct a narrative understanding of the world biased towards preferences. It is preferences which characterise the niche of the organism, such that it will act in order to satisfy this biased world view with suitable observations.

2.3 Operationalising Urban Liveability

Current operationalisations view urban liveability from a static perspective, invoking the existence of static dependent and independent variables to be measured and related to each other, as illustrated in Figure 4. To understand this approach, starting with a brief history of views on operationalising urban liveability and focusing on practices in the Netherlands is helpful. Valk & Musterd (1998) initiated the static approach by interpreting the fit between resident and environment as the valuation thereof, concentrating on satisfaction with life in relation to the living environment. Veenhoven (2000) noted that researchers had been slow to acknowledge the unavoidability of valuation as a measure of well-being. Furthermore, he posited that there are two operationalisations: apparent and perceived. According to Veenhoven, apparent liveability can only be measured through outcomes such as a healthy lifespan at the end of life. On the other hand, perceived liveability is to be studied throughout life and is measured using the valuation of a resident's living environment.

2.3.1 Well-Being across Resolutions, Domains and Time

Pacione (1990) proposed a modelling framework for well-being which incorporates both objective (environmental quality) and subjective indicators (urban liveability). He emphasized that subjectivity is about the behaviour-related function of interaction with the environment (fit). Figure 5 illustrates the need for objective and subjective indicators measured at different granularities and spatial scales. As is common in the static view of liveability, one can differentiate by attributes of the resident, such as class and age, as these can be measured in neighbourhood composition or survey respondents.

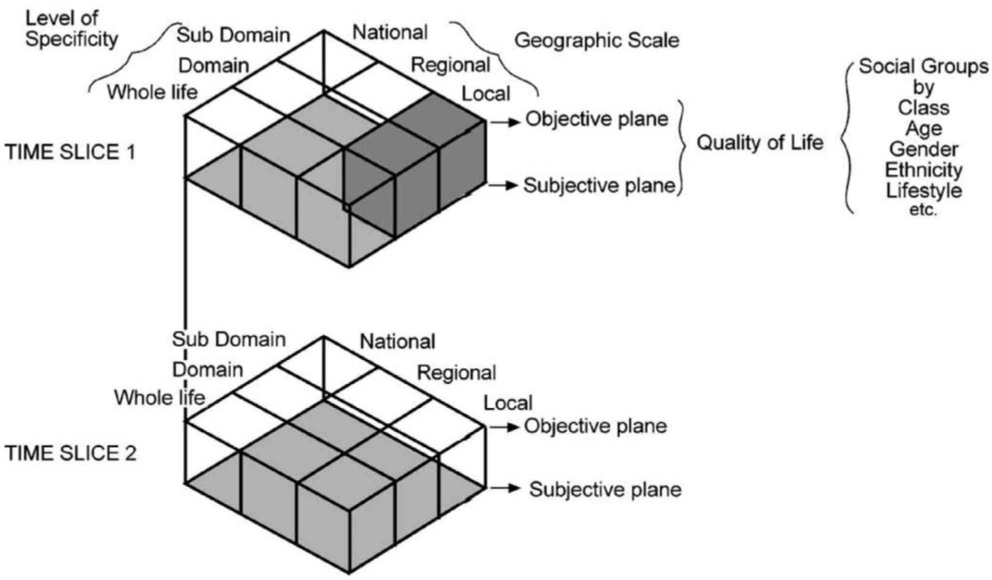


Fig. 1. A five-dimensional structure for quality of life research.

Figure 5: A five-dimensional model for quality of life research. From (Pacione, 1990).

Miller et al. (2013) addressed the problem of geographic scale, presenting a transport planning-focused operationalisation of liveability. After developing a comprehensive set of indicators relevant to liveability from a transportation perspective, local scales are considered by incorporating stakeholder perspectives in the multi-criteria analysis, which is a means to improve participation (Haezendonck, 2007). Lastly, Miller emphasised the geospatial nature of the transportation system's impact on liveability, providing a three-piece argument based on land use, space-time (congestion during rush hour), and spatial segregation of demographics. Additionally, externalities such as pollution or congestion are correlated across space, so spatial proximity explains a lot of variance. The resolution of analysis is therefore found to be relevant to expressing impacts on liveability.

2.3.2 Current Practise

Dorst (2005) outlined three operationalisations focusing on the role of indicators and residents' valuation of the living environment: perceived liveability, indicated liveability, and apparent liveability. Perceived liveability relies solely on residents' valuations, which may be stated or revealed. Stated preferences are acquired through surveys or interviews, while revealed preferences are patterns found in residents' collective behaviour, for example, travel behaviour or house prices. Indicated liveability, on the other hand, relies on a normative judgment by the analyst to determine what would likely make for a more liveable urban environment. Many operationalisations rely on this form in practice due to the ease of development. However, simplifying the burden of operationalisation on the analyst's side may have a significantly reduced validity, given the degree to which literature seems to indicate the role of perception or subjectivity.

Lastly, apparent liveability 'emerges' from the interaction between indicators and valuation, as proposed in Figure 6. For example, it can be done by estimating a regression model between house prices and indicators. The Leefbaarometer is an example of apparent liveability; it combines indicators and valuation through hedonic pricing and surveys (Mandemakers et al., 2021).

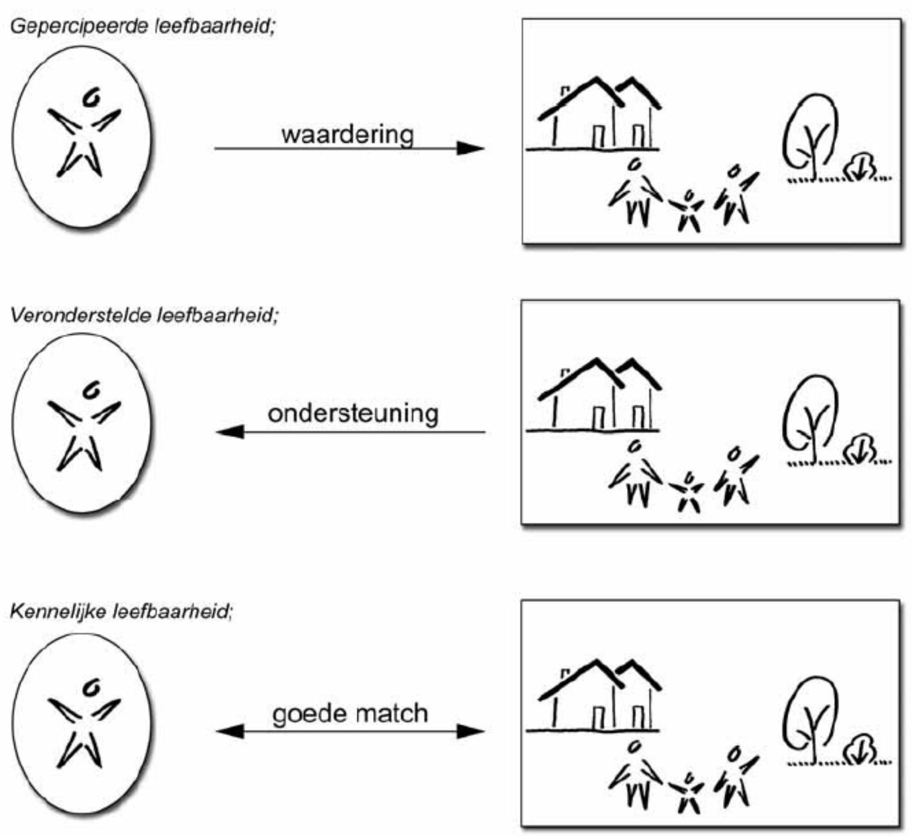


Figure 6: Three forms of operationalisation are used in current practice. Figure from (Dorst, 2005).

2.3.3 Three Facets of Subjectivity

Liveability is inherently subjective. Which environment promotes the greatest fit differs per individual and throughout their life. Operationalising the fit involves characterising the environment and its residents. Figure 7 provides a comprehensive overview of the fit of the resident environment, including characteristics of the resident, the environment, and their intersection.

Describing the environment involves indicators assigned to a spatial unit in an urban region. For example, large municipalities in the Netherlands have a liveability schema like 'leefbaarheidscirkel' in The Hague or 'wijkprofiel' in Rotterdam. Each neighbourhood is scored on various aspects, such as the availability of amenities and the number of robberies.

More generally, the environment is concerned with the built and natural environment and its resources, as presented in Figure 7. Residents' characteristics relate to their needs/desires, which exist to be satisfied. Lifestyles, culture, health, and personal characteristics describe needs/desires. They can also be used to provide practical segmentation into population clusters. For example, people of similar age brackets will have similar needs/desires in other aspects. Lastly, the perception of the environment separates liveability from environmental quality, which is deemed objective (Leidelmeijer, 2004). Additionally, perception relates to whether residents value quality of life and liveability.

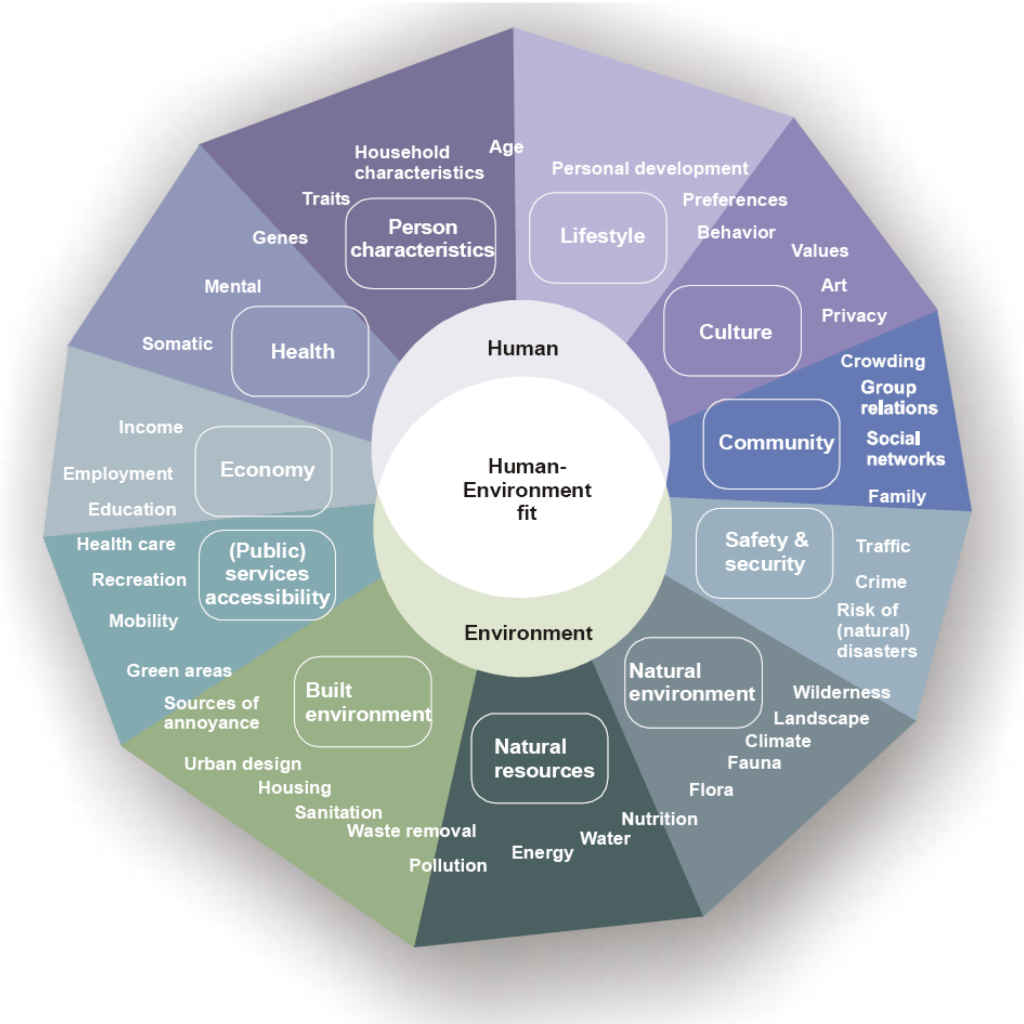


Figure 7: Overview of relevant aspects from literature as they relate to quality of life and liveability. From (Leidelmeijer, 2004).

Consolidating these three components into Figure 8 immediately clarifies where fit takes place. The conceptual model shows fit as a bidirectional relationship between percepts, the outcome of perception, and needs/desires. We will address the three components in reverse order, starting with needs/desires, perception, and indicators. It is this order that operationalisations seem to take when justifying the choice of indicators describing the living environment. Residents care about and value them, and they are perceived in daily life through experience.

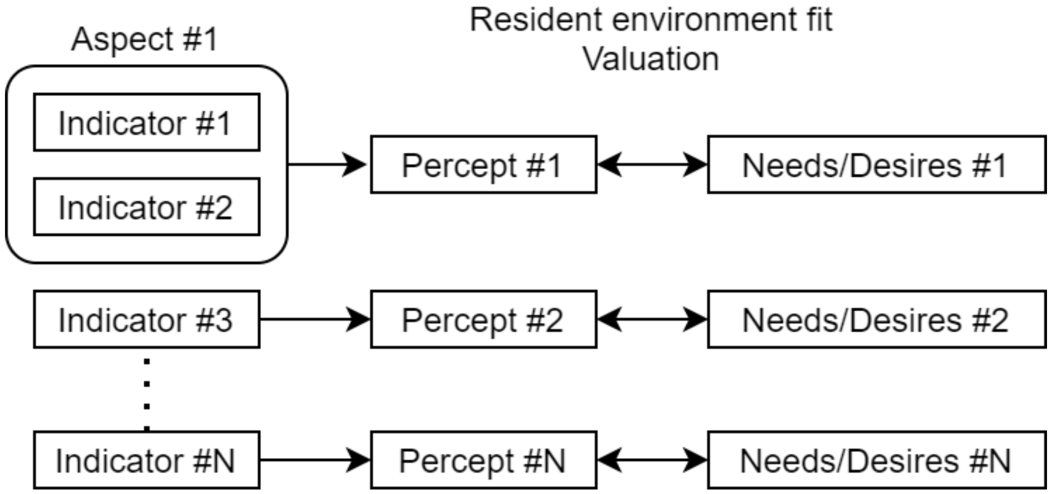


Figure 8: Conceptual model of the static approach to liveability.

Facet of subjectivity 1: needs & desires

Needs relate to basic requirements such as food and shelter. Desires to culturally embedded and personally relevant wishes (Veenhoven, 2000). The conceptual model in Figure 8 illustrates that needs and desires should be explicitly connected to indicators when crafting an operationalisation. Researchers often select indicators based on the broader literature. For example, Higgs et al. (2019) included a "daily living score" (measuring access to essential amenities) as an indicator of neighbourhood walkability, aligning with research from Marmot et al. (2010), which emphasises the importance of daily physical activity for health and lifespan. The need or desire in this example is health and lifespan, and since it is a form of indicated liveability, the analyst assumes that residents will assign a valuation. The importance of linking needs to indicators echoes the findings of many studies that refer to Maslow's hierarchy (Maslow, 1943). By drawing upon an established framework for defining needs and desires, analysts can more quickly define indicators without defining or studying needs and desires themselves.

Additionally, according to Veenhoven (2000), needs are measured through mood, whereas desires are measured through valuation. The argument is that needs are based on satisfaction, whereas desires are based on a wish. As such, the latter is quantified as the difference between the wish and the perception of the world. Needs are functional, contrasting wishes, which can be detrimental (Veenhoven, 2000). Finally, not every population segment will have similar needs and desires, nor will these be stable across an individual's lifespan. Hence, the overview in Figure 7 is relevant by noting the diversity of human attributes, going beyond just personal characteristics and encompassing lifestyle, culture, and health.

However, it is essential to note that liveability is a perspectival measurement from the perspective of the human towards the environment, as shown in Figure 2. Quality of life goes the other way. Hence, the static approach to liveability (with perspectival measurement) considers environmental aspects segmented by human aspects. The human characteristics are not the object of interest but a lens through which the measurement can be delineated into smaller population clusters. Aspects at the intersection of humans and the environment are more difficult to capture due to their complexity, e.g. social networks or health care and tend to involve subjective indicators.

Facet of subjectivity 2: perception

The perception of the environment makes liveability distinct from environmental quality, which is deemed objective (Leidelmeijer, 2004). When talking about different aspects of the environment, liveability tends to concern itself with the perception thereof by residents. When perception is not explicitly included in operationalisations, such as with indicated liveability, it is left implicit. That is, analysts assume that what their chosen indicators point to is perceived unambiguously.

The implicitness of perception may be of great value and justified for two reasons relating to population sampling. First, all other effects attributable to quality of life are cancelled so that every valuation points only to the presented indicators. Additionally, indicators are understood similarly due to, for example, having a similar culture. It is impossible to know if the analyst and respondent are talking about the same latent factor, the percept to which the indicator points.

The perspectival measurement involved in the static approach means that one measures either quality of life or liveability depending on the directionality. It is not easy to ascertain to what extent a valuation should be assigned to environmental characteristics or personal circumstances. As Veenhoven (2004) points out, the elated mood of the resident during the survey may be related to other parts of life besides the living environment, such as employment or social relationships.

However, as Leidelmeijer (2004) notes, not all aspects of the environment are perceivable. For example, residents cannot detect soil, water, or air pollution, but it still harms their health. Therefore, the valuation of these environmental characteristics is up to policymakers, who often set limits to concentrations of pollutants to prevent accumulation.

On the other hand, some studies go as far as to exclude objective environmental attributes altogether, focusing entirely on residents' valuation of the environment (Oviedo et al., 2022). This approach aligns with perceived liveability as defined by Dorst. Perceived liveability does not exclusively imply valuation/satisfaction of the living environment on the whole but should include delineation into indicators (Veenhoven, 2000).

In practice, there are no studies which make perception explicit. Instead, indicators or valuation are directly used. To highlight this implicit nature in current static operationalisations, Figure 9 depicts perception with a dashed box.

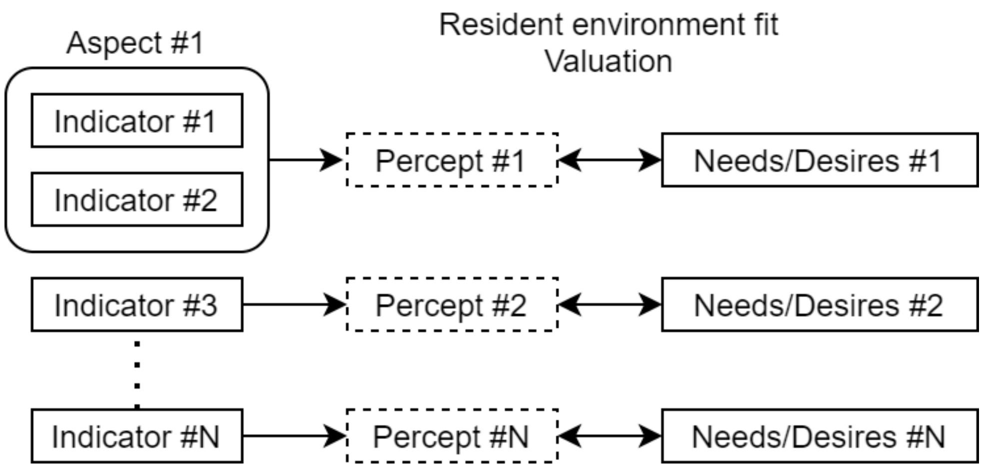


Figure 9: Conceptual model of the static approach to liveability, focussing on the implicit nature of perception in current operationalisations.

Facet of subjectivity 3: indicators

Having a combination of subjective and objective indicators is good practice. As Pacione (2003) states, *'we must consider both the city on the ground and the city in the mind'*. Other authors attest to this view (Leidelmeijer, 2004; Mandemakers et al., 2021).

As crafted by analysts, indicators represent the perceivable aspects of the world that relate to needs/desires. Dorst (2005) describes our ecological relationship with the built, natural, and social environments, each involving reciprocal interactions across several spheres. The analyst's task is to capture these elements with which residents have an ecological relationship, forming the indicators in indicated or apparent models of static liveability.

A prime example of this approach is the Leefbaarometer (Mandemakers et al., 2021). This state-of-the-art operationalisation defines five aspects of liveability: Safety, Social Cohesion, Housing Stock, Amenities, and Physical Environment. Each of these aspects is measured using a combination of objective and subjective indicators, providing a comprehensive assessment of urban liveability. See the following list to get a feel for the kinds of indicators comprising aspects.

- Safety: Measured using objective indicators like the number of violent crimes, vandalism incidents, and public disturbances, as well as subjective indicators such as survey results on the experience of safety and nuisance.
- Social Cohesion: Operationalized through subjective indicators like survey results on the experience of social cohesion and objective indicators such as the diversity of life stages in a neighbourhood and the population turnover rate.
- Housing Stock: Includes objective indicators like the vacancy rate of homes, the share of homes with a building height greater than 30 meters within 300 meters, the average size of homes, and the share of monumental homes and overcrowding.
- Amenities: Encompasses objective indicators like the distance to various services (schools, hospitals, cultural institutions, jobs), transportation stops, and job accessibility. The distance to amenities is often weighted using logarithmic functions to account for the diminishing relevance of amenities due to distance.
- Physical Environment: Includes objective indicators such as greenery and water, the risk of flooding and earthquakes, noise pollution levels, the amount of non-ionizing radiation, and how hot it feels on a hot summer day.

2.3.4 Automation of Perception

The resource-intensive nature of collecting indicators for comprehensive models has spurred the development of automated approaches. Remote sensing using aerial imagery has been employed to predict Leefbaarometer scores (Levering, Marcos, Van Vliet, et al., 2023). This approach shows promise in predicting aspects related to physical characteristics and housing stock, though it performs less well for features recognisable by proxy, particularly amenities. Other research has explored the contribution of various geospatial data to Leefbaarometer scores using inverse classification (Peeters, 2022), aiming to identify which actionable neighbourhood attributes most impact these scores. While advanced statistical modelling offers potential for longitudinal and explainable evaluation of liveability at high resolution, it is crucial to remember that the Leefbaarometer is a measurement model, not a causal one. It is intended as a signalling instrument to notify decision-makers of a need for further study, not as a predictive tool for liveability (Mandemakers et al., 2021).

Beyond data processing, deep learning models also serve to more closely approximate human perception (Dubey et al., 2016). Unlike the crafted indicators of analysts, people perceive their environment through images (Fan et al., 2023; Porzi et al., 2015; F. Zhang et al., 2018), smells (Jana, 2021), sounds (Gontier, 2021), heat (Hass et al., 2021), and (embodied) movement. Recent advances have focused on using street-view images to describe environmental qualities such as safety, beauty, wealth, and liveliness (Zhang et al., 2018). Others have included street-view images to operationalise the utility people experience in house relocation (van Cranenburgh & Garrido-Valenzuela, 2023). Reflecting on the conceptual model in these deep learning approaches simulates perception as a one-way process, treating it as an outcome (Figure 10). They consider the original data as an indicator and the resulting embeddings as perceptions. While this makes perception more explicit than implicit approaches, it still represents a static view of liveability and action perception. We will explore this argument further in the next section on ecological liveability.

A nuanced implication of using neural networks to approximate perception is the role of ambiguity. The conceptual models outlined so far all have a one-to-one mapping of indicators towards their associated percepts and needs/desires. However, introducing something that explicates perception means that the mapping between indicators and perceptions may not be one-to-one anymore. Indicators in this setting are high-dimensional rather than limited to traditionally understood indicators. So, there may be mischaracterisations between data and embeddings. For example, when a cat is classified as a dog. The loss of one-to-one mappings is due to the encoder's shared weights amongst different indicators.

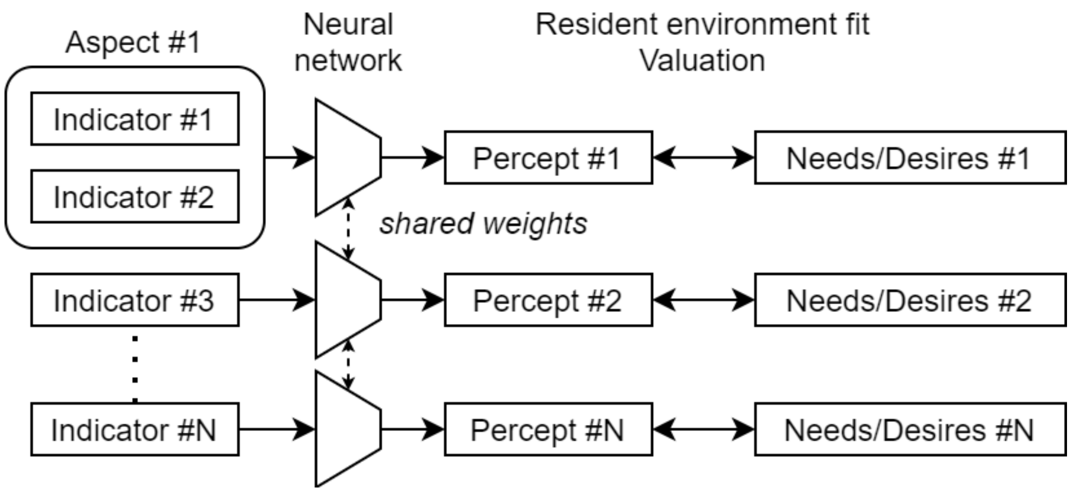


Figure 10: Conceptual model of the static approach to liveability focussing on the potential of neural networks to make perception explicit.

2.3.5 Process versus Content

Most efforts in operationalising liveability focus on content - the inventory of indicators and their associated valuations - rather than process, which involves how these indicators are selected and valuations determined. Ideally, indicators should align with needs and desires, but this link is typically drawn from literature rather than participatory methods. Analysts often presuppose the connection between needs/desires and indicators before attempting to incorporate perceptions.

This approach means that all operationalisations, not just indicated liveability, involve assumed relationships between indicators and needs/desires. These assumptions are implicit in indicators and valuations - what we ask about satisfaction inherently assumes its importance. Moreover, the methods for gathering valuations come from stated and revealed preference studies, which, while suitable for econometric analysis in narrow prosperity, may not fully capture the subjectivity central to liveability, as opposed to objective, rational economic choices.

Alternatively, the dynamic approach to liveability is process-oriented, as outlined in the next section. How indicators and their associated needs/desires are selected is made explicit in the generative model and the additional neural network, engine and transmission, respectively. The configuration of those parts should be informed by methods such as participatory value evaluation (Dekker et al., 2019; Mouter et al., 2021) to capture deontic values (de Boer et al., 2023).

2.4 Ecological Liveability

Ecological liveability represents an ecological perspective in understanding the relationship between residents and their environment. This section explores literature from various fields that share the idea that liveability is a dynamic process rather than a static outcome. Historically rooted in ecological psychology, these concepts are more familiar to architects than transport planners. Architects and urbanists have long focused on a sense of place (Appleyard, 1987), demonstrating a deep understanding of ecological psychology. The core of these ideas is that the city makes the places, not vice versa (Hillier, 2004). Places are moments, interactions with the physical urban environment.

Ecologically, niches (places) exist within habitats (cities). It does not make sense for a physical habitat to emerge from reciprocal relations unless one subscribes to idealism. Though neo-materialism nuances this problem (Rahmjoo & Albarracin, 2023). The directionality between habitat and niche is crucial for understanding the dynamic nature of liveability. As defined earlier, liveability relates to the fit between the resident and the living environment. However, how this fit is conceptualized can be static or dynamic. While previous chapters focused on static operationalizations, this section fully acknowledges the dynamic approach. This approach assumes perception as an action rather than an outcome, aligning with ecological psychology, transactionalism, and enactivism.

A recurring theme throughout this thesis has been the shift towards optimising well-being in policy evaluation through liveability and quality of life. This section will demonstrate how well-being, related to liveability and quality of life, can be operationalised through indicators and valuation and by crafting embodied dynamic models (Smith et al., 2022).

2.4.1 Behavioural Transport Geography

Behavioural geography covers many theories on the relationship between residents and their environment. Ecological psychology is of great value in behavioural geography due to its unique perspective on this relationship. Within transportation policy, ecological psychology provides theoretical backing for the link between travel behaviour and spatial, socio-economic, and personal characteristics (Acker & Witlox, 2008). A prime example is residential self-selection, which explains much of the variance in modelling the relationship between the built environment and travel behaviour (Kroesen, 2019; Van Wee, 2009). This concept highlights that people often live in areas that match their preferred travel modes rather than the environment influencing their behaviour. People who prefer to take the train and use active modes are more likely to live in cities. Vice versa, those who prefer to take the car tend to move to the suburbs to resolve the dissonance between preferred and actual travel behaviour (De Vos et al., 2012). The municipality of The Hague studied attitudes towards mobility, finding that clusters of attitudes align with the availability of infrastructure; those who like cars live in areas with low availability of public transport and vice versa (Coffeng, 2018). In sum, the characteristics of the environment do not influence behaviour insofar as people have already moved to areas that match them, resolving their travel behaviour dissonance.

Ecological psychology emerged as a response to the limitations of behaviourism, which focuses solely on the relationship between stimuli and behaviour (Moore, 2011). It recognises that behaviour is influenced not just by stimuli but also by desires and expectations, a wish for something in the world (Segundo-Ortin & Raja, 2024). In ecological psychology, perception and response are inextricably linked, with decision-makers perceiving options for actions or affordances (Rietveld & Kiverstein, 2014). Indeed, top-down action, selected from a set of affordances, precedes bottom-up input, the mixing of which forms perception. This non-trivial description of perception, beyond bottom-up processing, requires attention and will, therefore, be the focus of the remainder of this section, as well as that on mental representations.

2.4.2 Formalising the Dynamic Approach

The free energy principle and its corollary, the theory of active inference, provide a framework for operationalising action-perception loops across scales (M. J. D. Ramstead et al., 2019). This principle describes how self-organising systems stay far from thermodynamic equilibrium, explaining what stuff should do if it does not want to dissipate (K. Friston et al., 2014). Applied to living systems, it introduces the concept of characteristic states, or niches, which maintain the system's distinct existence (Bruineberg et al., 2018). Things emerge when stuff maintains characteristic states—a pullback attractor, as studied in dynamical systems (think whirlwind). The free energy principle provides the mathematics of 'things' or particles. A subset of these cognitive particles can select actions from a set of affordances. Another subset has temporal thickness provided by hierarchical structure (K. Friston et al., 2023). To understand temporal thickness, think about the pointers in a clock; the outside moves faster than the inside. The larger the radius, the more thickness. Those things without active states are like rocks. Those things with temporal thickness have a hierarchical structure and can account for longer spatio-temporal scales (sustainability).

An intuitive description of this dynamic process, which constructs and maintains niches, is to consider a room filled with soap bubbles—idealised particular things. Only those bubbles that have nestled themselves into a niche will stick around. Here, a niche is a timely balance between external and internal bubbles such that their pressures cancel each other out. The remainder of the bubbles which fail to satisfy this synchronicity will pop. Importantly, the surfaces never touch each other but only interact to push air currents, which apply pressure if concentrated. Sparse coupling entails that internal and external bubbles are never in direct contact but are always mediated by the bubble's surface of interest.

Furthermore, the internal dynamics of the bubble have characteristic states such that it will gradually wiggle towards a setting in which it does not collide with external bubbles without timely internal compensation. A pullback attractor of internal bubbles fulfils the role of preferences, so the bubble of interest will act accordingly; if it does not, it dies. The wiggle is more formally known as Langevin dynamics, which describes a deterministic path with stochastic noise. The free energy principle boils down to the observation that whatever is left out there in the world to be observed should probably follow this principle; otherwise, it would have dissipated (bubble pops). In turn, the free energy principle is somewhat deflationary, like a tautology, as Friston acknowledges (K. Friston, 2018).

Active inference applies the free energy principle by describing what these systems do. Where the free energy principle is a principle, active inference describes its dynamics, and Bayesian mechanics elaborates on its mechanics (M. J. D. Ramstead et al., 2023). The essence of all these descriptions is Lagrangian equations, which operationalise the concept of paths of least action; the most likely evolution of states throughout time is the one that deviates the least from an optimal path, as deviation requires action. It is subsequently a computational problem to approximate this optimal path, as it is not analytically derivable for real-world problems.

While a comprehensive technical treatment of either the dynamics or mechanics is beyond this thesis's scope, it is useful to delineate some key concepts to understand the concept of niche. Active inference posits decision-makers as being conditionally independent of their environment by introducing a Markov blanket, Figure 11. This Markov blanket separates internal and external states, with the dynamically coupled system aiming to balance these states by employing active and sensory states, Figure 12. More accurately, the relationship is sparsely coupled. Such that internal states do not affect external states directly and vice versa.

The minimisation of differences between internal and external states is mediated by action and perception, represented by active and sensory states. A key component of active inference is partial observability: decision-makers never have access to the 'true' world, only their observations (and actions). They constantly attempt to infer the causes of their observations, which are the hidden states of their world model. Furthermore, self-evidencing is when the cause of observation is one’s action. This inference of latent causes from observations makes active inference a Bayesian approach to perception, Figure 13.

Predicting observations, known as predictive processing (Clark, 2013), leads to efficiencies as only divergences between top-down predictions and bottom-up observations must be processed. This constant flux of errors continuously updates the world model throughout life, unlike neural networks, which rely on backpropagation (Millidge et al., 2022).

Active inference can be framed as a study of generative models. These models are joint probability distributions that contain active states in addition to observations in the case of enactive models. The generative model of inferring hidden states from observations, as illustrated in Figure 13, is as follows:

$$P(s, o) = P(s)P(o|s)$$

Marginalisation of this generative model allows for the calculation of individual probabilities, e.g. the probability of an observation:

$$P(o) = \sum_{s} P(s, o)$$

Bayesian updating is how the generative models are improved by incorporating new observations/evidence. Posterior distributions result from such an update. An example is presented below. A marginalisation term is added in the numerator to ensure a total probability over states of 1:

$$P(s|o) = \frac{P(o|s) * P(s)}{P(o)} = posterior = \frac{evidence * prior}{marginalisation}$$

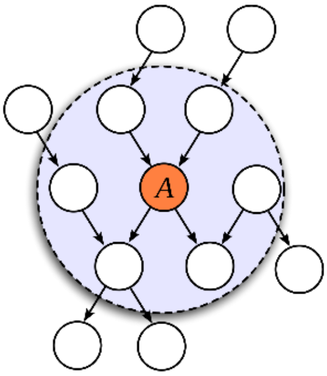


Figure 11: Illustration of a Markov blanket in a Bayesian network. Circles contain probability distributions and arrows, which are their interdependencies. One can write the probability of A given its parents. The dotted line denotes the blanket. From Wikipedia.

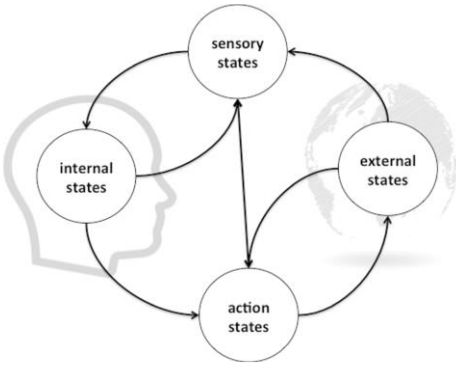


Figure 12: reciprocal relationship between agent and environment. Markov blanket returns through sensory- & active states. Sparse coupling via blanket ensures the agent is distinct but ecologically coupled to the environment. From (Sims & Pezzulo, 2021).

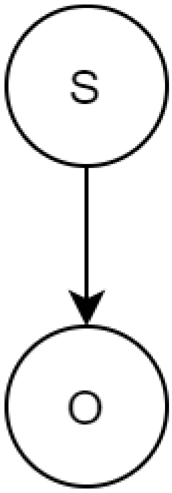


Figure 13: Causal relationship between hidden states and observations.

Action in generative models is incorporated by adding policies to the generative model. Policies are sequences of action that apply to the hidden states. Observations depend on hidden states and policies, whereas hidden states depend on policies, and policies are independent of the others.

$$P(o, s, \pi) = P(o|s, \pi)P(s|\pi)P(\pi)$$

A more complete formalism of such generative models can be found in Figure 14. Hidden states progress, while observations are inferred through hidden states, and action applies to hidden states at the transitions between time steps. At this point, it is worthwhile to reintroduce the bubble metaphor; internal bubbles push against the main surface until external ones push them back again. The internal states palpate the blanket in expectation of colliding with an external bubble/state. The divergence, anything but equal pressure of bubbles, is propagated throughout the model's parameters for updates. While these models run on computers and cannot die if they fail to predict well enough, alternative directions propose the need for mortal computation (Ororbia & Friston, 2023).

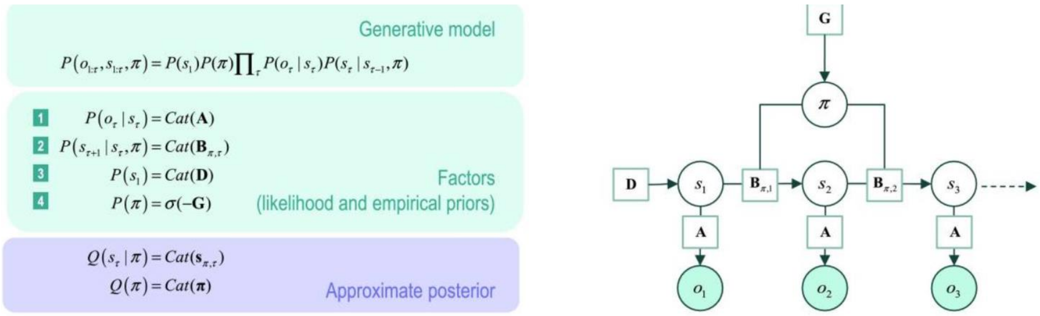


Figure 14: Generative model for discrete (categorical—Cat) states. From (K. J. Friston et al., 2017).

In contrast to the static perspective of liveability, which considers fit between resident and environment to be a valuation, the dynamic approach posits that fit occurs through perception. Panel A in Figure 15 illustrates how the static approach locates fit between percepts and needs/desires. In contrast, the dynamic approach proposes that fit occurs between indicators and percepts—observations and hidden states. Panel B illustrates how neural networks can map high-dimensional data to lower-dimensional hidden states (image to a few numbers). Individual images are indicators; their representations are hidden states. Moreover, panel C unrolls the dynamic approach presented in A out over time.

In line with active inference, policies impact the world model and tend to align with needs and desires. Policy selection is based on expected free energy (G). It is calculated by imagining actions and their resultant observations using an approximate generative world model: the variational density Q. Only affordances (choice set) can become sampled policies. Expected free energy is a combination of information gain and pragmatic value:

$$G(\pi) = Epistemic\ value + Pragmatic\ value$$

The discrepancy between the world model and observations represents the fit between the resident and the environment. More formally, this discrepancy is variational free energy—minimising it implies perceiving observations such that the percepts align themselves to be as simple (low complexity) and accurate as possible, like Occams razor:

$$F(Q, o) = Complexity - Accuracy$$

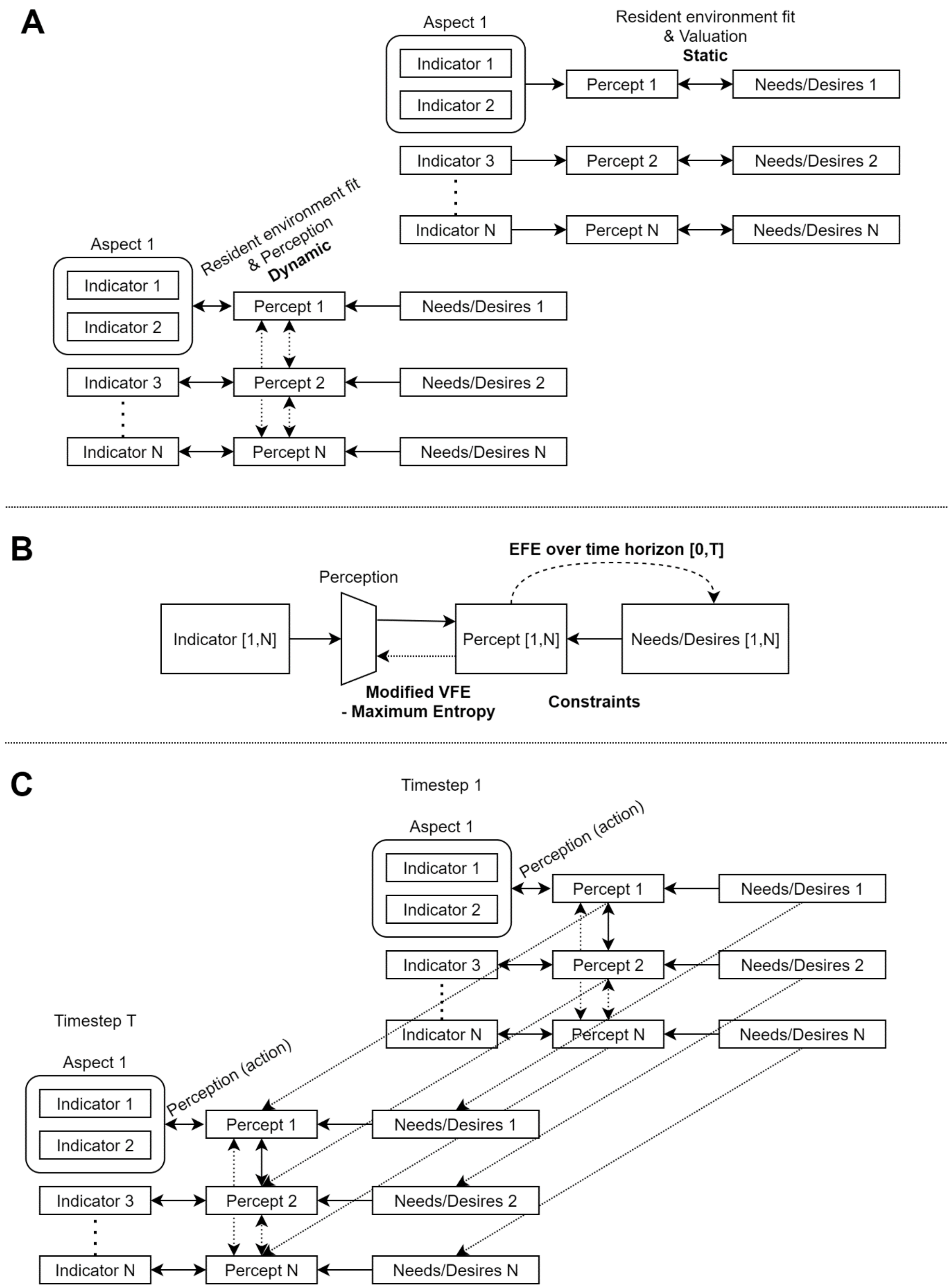


Figure 15: The place of fit between environment and resident defines the difference between a static or dynamic interpretation of action perception. Figure ***A*** *shows two models; the top right illustrates the static approach, and the bottom the dynamic. Figure* ***B*** *builds upon the dynamic approach, showing an active inference implementation where a neural network is inserted into the bidirectional relationship—the neural network is akin to the transmission in an engine. Figure* ***C*** *expands upon* ***A*** *by adding temporal depth. The initialisation, t = 0, is left out for visualisation purposes.*

2.4.3 Affording Action Selection

The role of affordances is central to the dynamic approach to liveability. In ecological psychology, perception is only possible through action. Gibson (1986) developed the notion of affordances: "*The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill.*" Similarly, in active inference, action plays a role in perception, with Parr et al. (2022) referring to this as palpation: hearing is listening, and seeing is looking. Rather than perceiving raw sensory data in a bottom-up manner, the dynamic approach implies that the potential for action itself is sensed.

Activities studied in transport planning and places studied by urbanists can be unified under affordances. The function of cities is thus constructed through interaction distributed across its landscape, taking the term 'landscapes of affordances' proposed by Rietveld & Kiverstein (2014) quite literally. This interpretation is enabled by considering accessibility to activities afforded by the transportation network, where higher access to other areas is akin to hills in a landscape. See Figure 16 and Figure 17 for visualisation of location-based accessibility (building density and travel time) where warmer colours are at higher elevations. Just like one can walk up a hill to see far into valleys—one can go to a central train station or highway entrance to access many destinations.

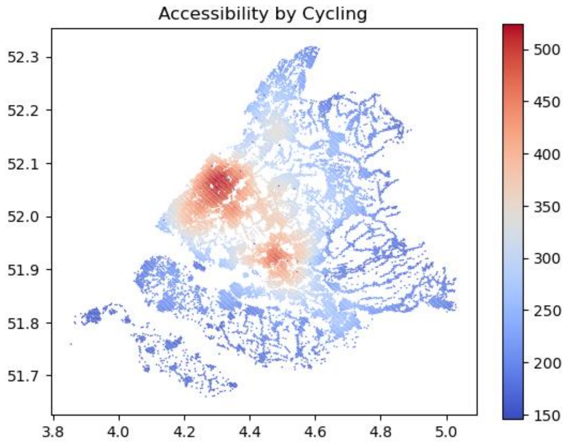


Figure 16: Location-based accessibility for cycling in the province of South Holland.

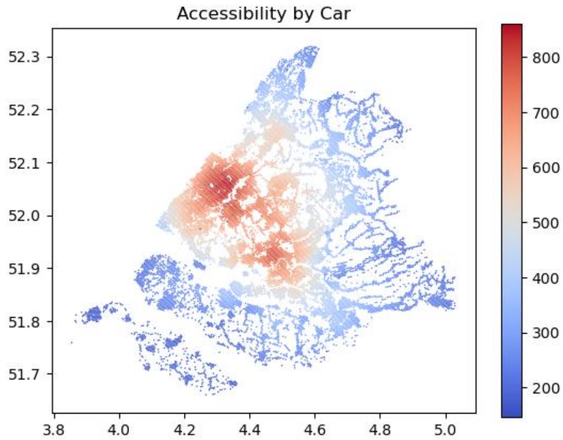


Figure 17: Location-based accessibility for driving by car in the province of South Holland.

In addition to the landscape of affordances, there are also fields of affordance (Bruineberg & Rietveld, 2014). As White & Miller (2024) put it: *"the landscape of affordances applies to those affordances that make up an agent's ecological niche, available to us through our socio-cultural practises. It is broad, encompassing all of the theoretically available affordances in my local environment. Right now, my affordance landscape is the city of Brighton, or rather, Brighton understood on the spatio-temporal scale of its particular stable patterns of shared, public, affordance."* Unlike landscapes, fields are unique to the individual and change as they interact with the world. It relates to those affordances that become relevant in the moment and are guided by attention, where attention is not much more than precision weighting in active inference, i.e., higher precision on policies that influence hidden states (Clark, 2013).

The demarcation between fields and landscapes of affordances enables smart ambient environments. See White & Miller (2024) for a synthesis thereof with active inference. Such smart ambient environments tune not affordances themselves, like infrastructure or objects, but the fields of affordance, the process which guides the perception of affordance. In turn, these ambient environments may improve allostatic control—the ability to take suitable actions in expectation of the future—which leads to a better grip on the environment. The grip on one's environment is a central concept in active inference; it is the purpose of taking actions and a description of how capable those actions are of influencing the environment as required by needs/desires. If one has a low grip and continues to do so, then mood may lower (Bruineberg & Rietveld, 2014).

Practically, ambient smart environments open or close the field of affordances, such as nudging agents to explore or exploit more. Exploration may be helpful when the choice set is larger than considered, e.g. habitual car drivers who could use public transport (given that these options have similar utility). Exploitation, on the other hand, requires narrowing the field of affordances. What is paid attention to should be constricted, enabling focus on a narrow task of interest. As Bruineberg and Rietveld discuss, OCD is an extreme example of this narrowing, as singular affordances stand far beyond the rest of the field.

2.4.4 Niche Construction

Niches and habitats build on affordances by demarcating physical and lived environments. Revisiting the active inference formalism of partitioning into states, Figure 12, internal states aim to approximate external states, becoming a mirror image inferred through observations. The causal mapping inference process is optimised by minimising free energy, which is the continuous fitting process. Since external states are inferred through observation, the physical environment, as studied in liveability, does not exist per se. Rather, the fit is between the resident and the niche (Bruineberg, Rietveld, et al., 2018). The physical urban environment, the city, may be understood as the habitat, while the places lived in are niches. The city creates a place through the unfolding of perception, in which action is the engine guided by affordances within the niche. Importantly, residents live in places or niches rather than physical environments or habitats. The niche contains cues to action (affordances), and these affordances are perceived rather than raw sensory bottom-up input. Residents may change the physical environment (habitat) to construct a shared niche or place, a process known as niche construction. Niche construction allows analysts to reframe the static interpretation of (urban) liveability as a dynamic process. Building upon cultural, cognitive, and affective niche construction, Constant et al. (2022) define it: "*Niche construction is the process through which organisms create and maintain cause-effect models of their niche as guides for fitness influencing behaviour.*"

In urban settings, niche construction takes concrete form in infrastructure and dwellings. While this process is largely formalised in developed countries, developing countries illustrate that niches will be constructed in any way possible, as seen in favelas (Downey, 2016). Lives in favelas are distinct from those in formalised neighbourhoods, often yielding much lower standards of living due to disease and stress. This intense stratification of urban regions leads to disparate distributions of stress. Nagatsu et al. (2023) aptly state, "*Cities are cognitive, technological, and cultural niches that have enabled unforeseen amounts of innovation, economic development, and technological evolution. In many respects, cities should be perceived as a pinnacle of niche construction, and cities can be fruitfully understood as ecosystems of ideas*."

Applying niche construction to urban settings, such as cycling in cities, reveals interesting dynamics. Nagatsu et al. (2023) identify four model assumptions for reasoning about niche construction: 1) humans inhabit niches which provide affordances, 2) affordances enable and constrain behaviours which also spread through social networks, 3) afforded behavioural patterns reinforce learned behaviours, and 4) affordances may be intentionally refined and altered. These assumptions lead to potential feedback loops: improved cycling infrastructure creates affordances promoting the uptake of cycling behaviour, behaviours are copied and spread, users improve their cycling and navigational skills, increasing use in the future, and new niches may be constructed as cycling uptake increases among the population. Empirical results support niche construction in terms of city cycling behaviour (Kaaronen & Rietveld, 2021; Kaaronen & Strelkovskii, 2020).

2.4.5 Niche Construction and Transportation Policy

Modern transportation policy, with indicators focussing on accessibility, liveability, and safety (Huibregtse, 2021), has slightly different interpretations under the dynamic approach to liveability, which explicitly concerns niche construction. Accessibility, in this context, is more than an indicator. Instead, accessibility relates to the landscape of affordances based on location-based accessibility, which accounts for the attractiveness or opportunities a location affords. This concept of affordances is central to niche construction; the choice set is the one that affords the action necessary to construct a niche. Accessibility is, in turn, mechanised in the dynamic approach to liveability. Furthermore, option value, as it relates to the benefit of having redundant affordances in one's living environment (K. T. Geurs et al., 2006), becomes explicit through the maximisation of information gain as the dynamic approach considers both pragmatic and epistemic value.

One example of niche construction is residential self-selection. Residential self-selection is the phenomenon in which travel behaviour is explained away through either geographic characteristics or personal travel preferences. Controlling for personal travel preferences leads to a zero loading on geographic characteristics in the structural equation model. Hence, residents who prefer certain travel affordances will have moved their house to a niche satisfying these. Self-selection is one of the three ways to maximise fit in the dynamic approach to liveability. 1) A resident can change the mapping of indicators to percepts, seeing the world differently. 2) They can act to change the world, for example, by buying a car. 3) Or they can move towards another urban region or neighbourhood.

As described under the broad prosperity framework, capabilities are a relevant policy consideration, as physical infrastructure by itself is useless if it cannot be used (Snellen & Bastiaanssen, 2021). Affordances are practically equivalent to capabilities. Equipped with the necessary capabilities, residents can actively explore and shape their urban environments, constructing niches that maximise fit. These niches extend beyond the physical characteristics of the living environment, encompassing social connections and exchanging ideas (Constant et al., 2022). As Pentland (2020) describes, this' idea flow' fuels economic growth and innovation, flowing most freely across 'social bridges' that connect disparate social clusters. Lucas Spierenburg et al. (2023) have developed a method to quantify social segregation in urban areas using three indicators: intensity separation and scale. First, using demographic variables assigned to each neighbourhood, agglomerative clustering is applied to create clusters of similar demographic composition. Then, indicators are calculated for these larger clusters of the urban region. Intensity and separation relate to the spatial distribution of groups across regions, whereas scale relates to the size of the segregated regions. Accessibility with walking as modalities gives the final exposure of each cluster.

Active inference models provide a framework for understanding and optimising this idea flow within communities (Albarracin et al., 2022; Catal et al., 2024). Social cohesion becomes an integral part of niche construction rather than an indicator to measure. Each interaction, whether with neighbours or visitors, contributes to this ongoing process of niche construction. Urban downtowns become the place for exchanging ideas, fueled by the daily influx of commuters from diverse backgrounds.

Cultural narratives may shape future transportation systems. Active inference models run on narratives (Bouizegarene et al., 2024). Dissemination of narratives, or the segregated manner thereof, has implications regarding resilience against hybrid warfare (Waltzman, 2017). The role of digital communication in niche construction warrants careful consideration. As described by the concept of triple access planning, digital communications are just as essential to transport planning as infrastructure and accessibility since they address the problem of moving somewhere else to perform an activity (Rye et al., 2024). Reiman (1995) underscores the need to balance connectivity with concerns about privacy and surveillance, avoiding the creation of a digital panopticon. The term panopticon is derived from the Alcratez prison, in which a single centralised observation tower looks at the entire prison. It describes people adjusting their behaviour with the knowledge that surveillance is everywhere, even if the guard is not looking at that moment.

2.5 Mental Representations

The Free Energy Principle (FEP) serves as a scale-free operationalisation of the action-perception loop, providing a unifying framework for understanding the fit between residents and their environment across different scales. This principle posits that systems aim to minimize free energy, equivalent to maximising model evidence—the probability of observations given the model (Parr et al., 2022). Dynamic systems modelled under the FEP perform inference over beliefs, maintaining internal models that are continuously updated through interaction with the environment (M. J. D. Ramstead et al., 2024). This process is framed as Bayesian inference, where systems maintain and update internal models based on their experiences. These internal representations, or parameters of the mathematical model, encode beliefs about states and are adjusted to guide future actions, reflecting an optimistically biased goal-directed behaviour in the action-perception loop. Within this framework, organisms (in our case, the resident) embody their niche. That is, the niche is the generative model comprised of internal and external states, which are given blanket states and are constantly updated.

Furthermore, the organism, being embodied and enactive, is its niche. That is, residents are places within spaces, creating places through niche construction. Where parameters encoding internal states track external states. The niche is formed through the joint generative density of internal, external, and blanket states, which are dynamically updated by following gradients of free energy (M. J. D. Ramstead et al., 2024). The FEP framework also provides insights into goal-directed behaviour and teleology, as systems minimizing expected free energy exhibit sophisticated goal-directedness. This goal-oriented understanding extends across various scales and complexities (Beni & Friston, 2024).

Figure 18 illustrates the circular causality between action and perception in active inference. Panel A shows how preferences shape prior percepts, which, when combined with indicators (evidence), become recognized posterior percepts. Panel B introduces the concept of policies (action sequences) and observations, demonstrating how top-down priors (actions) interact with bottom-up sensory information. Panel C presents a more detailed POMDP scheme, showing how action policies (π) are selected based on expected free energy (G) and how they influence state transitions over time.

The recognition density, representing the organism's 'best guess' about the causes of its sensations, emerges as a posterior belief after the interaction of top-down priors (actions) and bottom-up sensory information (M. J. D. Ramstead et al., 2020). This process highlights the enactive nature of perception in active inference, where the organism actively shapes its sensory inputs through action.

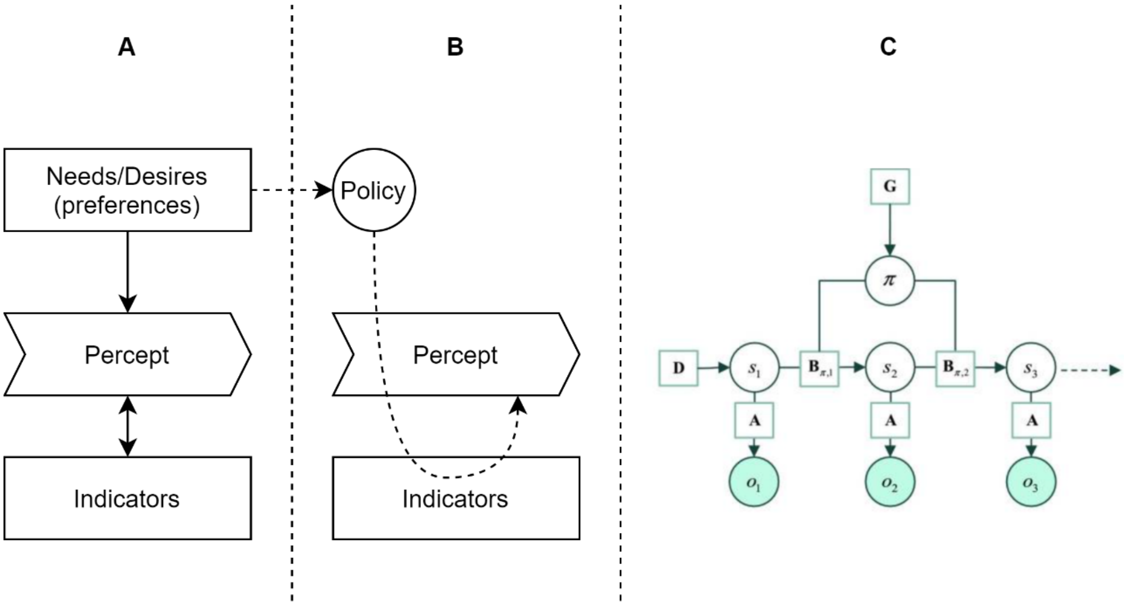


Figure 18: Schematic representation of active inference. Panels A & B are our work. Panel C is from (K. J. Friston et al., 2017). (A) shows preferences (needs & desires) optimistically influencing the unfolding world model, which generates perceptions. These perceptions then adjust the model. (B) introduces policies (action sequences) and observations, demonstrating top-down policy priors interacting with bottom-up sensory information. (C) illustrates a partially observable Markov decision process (POMDP) scheme, where policies (π) are selected based on expected free energy (G) and influence state transitions over time.

2.5.1 The Debate on Mental Representations in Philosophy of Mind

In the philosophy of mind, mental representations have long been a subject of intense debate, centred around two fundamental criteria: their capacity to describe the external world and their instrumental role in reasoning and action.

Representationalism posits that cognition fundamentally involves internal models, which can be either symbolic or encoded in neural networks. These models are seen as organized internal states, separate from the external world, that strive to mirror the structure of the environment and accurately describe it (Sims & Pezzulo, 2021). Proponents of this view argue that these internal representations are crucial for reasoning, decision-making, and action (Engel et al., 2016). They contend that the brain builds and manipulates these representations to make sense of the world and guide behaviour.

On the other side of the debate, non-representational views challenge the necessity of detailed internal models for cognition. Among these, enactivism has emerged as a prominent theory, emphasizing the direct coupling between an organism and its environment. Enactivists argue that cognition arises from this dynamic interplay rather than from internal models. They view internal states as less distinct from the environment, rejecting the need for them to mirror the world's structure. Instead, they focus on sensorimotor contingencies and the direct perception of affordances, highlighting the role of embodied action and interaction with the environment in shaping cognition.

The enactivist perspective suggests that the external world itself serves as the best model, negating the need for detailed internal representations (Bruineberg et al., 2018). This view aligns with theories of extended cognition, which propose that cognitive processes extend beyond the boundaries of the brain to include the body and environment (Constant et al., 2022). From this standpoint, cognition is seen as an ongoing, dynamic process of interaction with the world rather than a series of computations performed on internal representations.

As the debate between representationalism and non-representationalism has unfolded, new frameworks have emerged that seek to bridge the gap between these seemingly opposing views. Active inference, in particular, offers a potential reconciliation by incorporating elements of both representational and enactive approaches. This framework maintains the concept of internal models but frames them as dynamic, action-oriented control systems that guide adaptive behaviour. Active inference provides a formal understanding of how action and perception are intertwined, capturing both representational and non-representational processes (Constant et al., 2021).

Active inference suggests that cognition can involve rich, reconstructive internal models while also accommodating more direct, embodied interactions with the environment. This nuanced perspective acknowledges the role of representations in certain cognitive processes while recognizing the importance of immediate environmental coupling in others. This ongoing debate in the philosophy of mind mirrors broader discussions about the nature of cognition itself. As researchers continue to explore the complexities of mental processes, the distinction between representational and non-representational approaches becomes less of a theoretical exercise and more of a practical consideration in modelling cognitive systems, such as those required to simulate the dynamic approach to liveability. Exemplifying the practical application of this debate is the difference between static and dynamic approaches to urban livability. A purely representational view might focus on discrete choice models (van Cranenburgh & Garrido-Valenzuela, 2023), while a non-representational approach would emphasise the immediate, embodied experiences of urban spaces by using virtual or augmented reality to study the interaction between, for example, cars and pedestrians.

2.5.2 Representation Wars: Enacting an Armistice

Over the past three decades, the philosophy of mind has been marked by the "representation wars", with ongoing debates between representationalist and dynamicist positions. Representationalists argue that cognitive processes involve rich, reconstructive internal models, while dynamicists view cognition as arising from direct interactions with the environment without the need for detailed internal representations (Constant et al., 2021). Recent developments in active inference propose a way to reconcile these perspectives by showing that a niche's generative model can encompass both representational and non-representational processes. The usage of niche is critical; we are not talking about an isolated brain in a jar but an embodied enactive organism as described by theories of extended cognition (Constant et al., 2022). Active inference posits that the brain engages in a form of inference, using generative models to predict sensory inputs and guide actions. This process involves optimising beliefs about hidden states—the causes of observations—through embodied action. Representational pathways involve the manipulation of beliefs about hidden states, which is essential for tasks requiring detailed internal models. In contrast, dynamic pathways rely on direct sensorimotor contingencies and the immediate coupling of perception and action, which are characteristic of enactivist views.

In the representational pathway, detailed internal models of the world are updated and maintained. These models are represented by the generative model P(o,s,π) and the approximate posterior Q(s,π). The generative model encodes the agent's beliefs about how sensory observations are generated, while the approximate posterior represents the agent's best guess about hidden states and policies given sensory data.

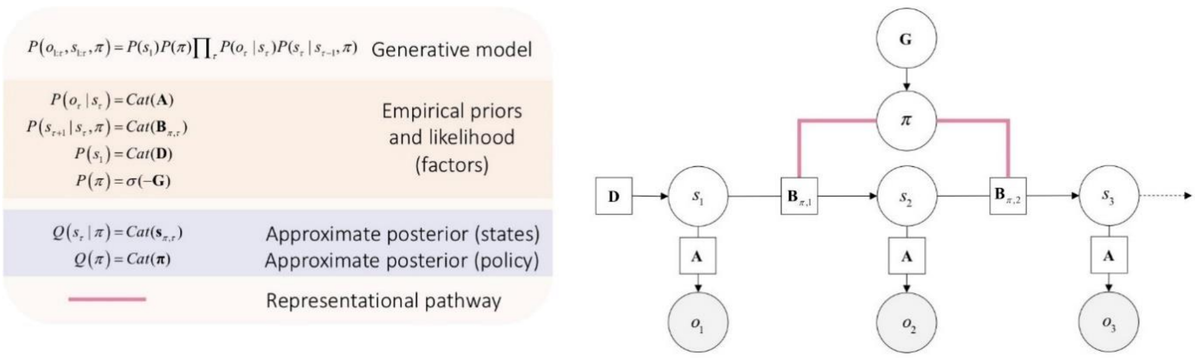


Figure 19: Representational pathways in active inference. The left side shows the mathematical formulation of the generative model, empirical priors, and likelihood factors. The right side illustrates the graphical model of how states, observations, and policies. From (Constant et al., 2021).

The dynamic pathway in active inference represents a more direct, embodied approach to cognition. Instead of relying on detailed internal models, this pathway leverages deontic value - a direct mapping from policies (action sequences) to expected observations. The dynamic pathway allows for fast, frugal decision-making based on learned associations between actions and their typical outcomes in the environment. Habitual action-perception loops can be particularly useful for well-learned behaviours or in situations requiring rapid responses.

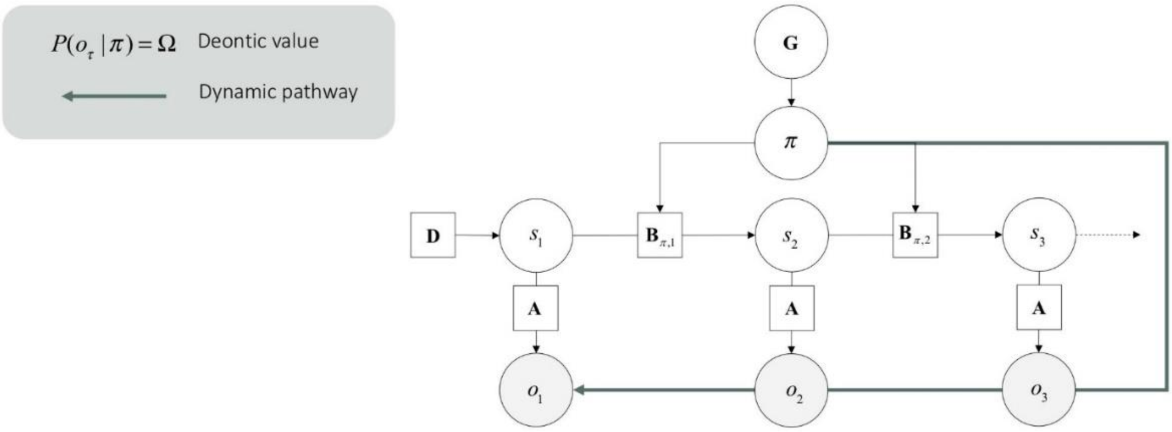


Figure 20: Dynamic pathways in active inference. This figure introduces the concept of deontic value P(o_τ|π) = Ω, which represents a direct mapping from policies to observations. From (Constant et al., 2021).

By incorporating representational and dynamic pathways, active inference provides a framework that accounts for a wide range of cognitive processes, from deliberative planning to intuitive, embodied actions. This dual-pathway approach offers a potential resolution to the "representation wars," acknowledging the value of both internal models and direct sensorimotor couplings in cognition. Further similarities may be drawn between system one and two thinking as proposed by Kahneman (2011), where system one is like the dynamic pathway, whereas system two is more like the representational pathway.

2.5.3 The Role of Representations in Action-Perception Loops

The action-perception loop is fundamental to understanding cognition, where action and perception are interdependent rather than sequential processes. Furthermore, this means that representations arise where there is no direct interaction with the environment; internal states are, therefore, representations while still fully subscribing to an enactive view of cognition aligned with the action-perception loop and ecological psychology (Bruineberg et al., 2018). This view challenges traditional models of cognition by emphasizing action-oriented perspectives, a pragmatic turn (Engel et al., 2016). Perception is not a static outcome but a biased dynamic process involving palpations, looking to see, listening to hear, etcetera (Parr et al., 2022).

Self-evidencing is a concept where organisms act to gather evidence for their existence. It involves actively shaping the environment to align with internal predictions, extending the self-evidencing concept to encompass niche construction (Constant et al., 2018). Context plays a crucial role in interpreting sensory information, with hidden states in generative models providing the necessary context for inference. The brain's task is to infer these hidden states from sensory data and prior knowledge, maintaining organization within the environment (Bruineberg et al., 2018). Context unfolds as residents go out in the world, and the context is about the flow of narratives of observations. Not discrete choice moments.

2.5.4 Integrating Predictive Processing and Enactivism

Predictive processing is a core conceptual pillar in active inference. The combined interaction of top-down predictions and bottom-up errors enables dynamic rather than static approaches to modelling cognition and liveability. Pezzulo et al. (2024) address the differences between static and dynamic approaches as seen in generative artificial intelligence, noting that embodiment and predictive processing, as provided for by enactivism, are distinct from current deep learning approaches. Figure 21 shows how each layer in predictive processing aims to predict its downstream neighbour such that error is calculated locally rather than globally and subsequently back-propagated.

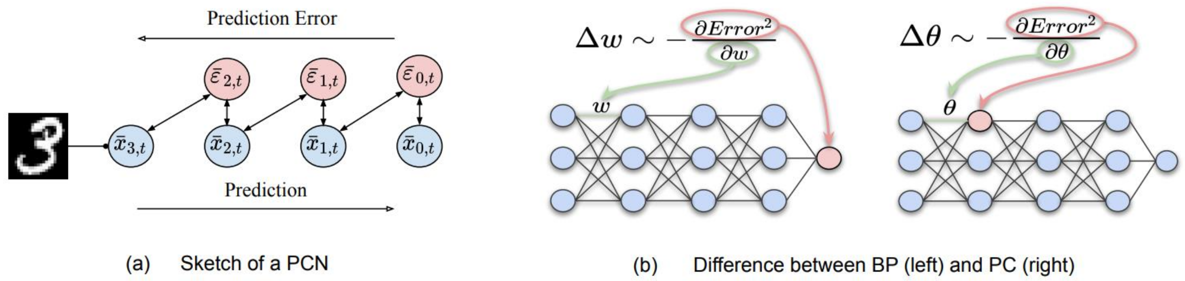


Figure 21: Conceptual overview of deep learning using backpropagation and error dynamics using predictive coding (network). From (Millidge et al., 2022).

From a predictive processing perspective, Hutto & Hipólito (2021) argue that perception is neither radically enactive nor purely representational but a hybrid of both. Predictive processing suggests that perception involves active hypothesis testing, where sensory input is continuously compared against predictions generated by internal models. This process incorporates both top-down and bottom-up information flows, blending elements of enactive interaction with representational structures. However, the persistence of perceptual illusions, like the Müller-Lyer illusion, challenges purely predictive models, suggesting that basic perception is habitual and non-inferential, while higher-order perceptual judgments are inferential (Hutto & Hipólito, 2021). Basically, some visual perception components have parameters that cannot be inferred but are fixed, like the dynamic pathway in representation wars.

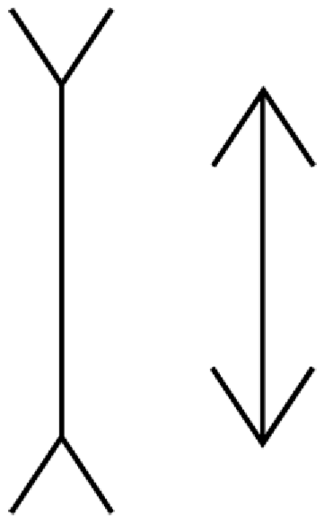


Figure 22: Müller-Lyer Illusion

Hence, it could be argued that representations studied in neural networks are distinct from those studied in the philosophy of mind. In connectionism, network weights are representations. In enactivism, mental representations are used to parameterise the action selection process. However, these representations are not always needed; some parts of action perception are non-inferential. All this seems straightforward until one takes note that input-output—deontic pathways—are like trained neural networks. Therefore, the representations in neural networks are not so much behaviourist as they are enactive. The difference found in the demarcation between training and inference is that neural networks using back propagation do not learn during inference. In contrast, predictive processing models learn throughout and during operation.

2.5.5 Practical Considerations

Challenges in dealing with high-dimensional data within generative models include managing computational complexity and ensuring efficient learning and inference. Active data selection, evaluated through the metric of information gain, is a crucial concept here. Information gain can be decomposed into ambiguity and predictive uncertainty, guiding the selection of the most informative data points to improve model accuracy (Parr et al., 2024).

Deep neural networks, particularly those aiming to simulate the recognition density, offer a powerful tool for handling high-dimensional data. These networks can efficiently process complex sensory inputs, learning intricate mappings between input data and internal representations (Mazzaglia et al., 2022). Different methodologies can be used to train these deep neural networks. Figure 23 below mostly involves the reconstruction of the original data, as illustrated by the hourglass shape of the red and blue neural networks. Alternatively, one could tune these representations to a task. In general, tuning neural networks with a task in mind leads to strong performance on that task, for example, by predicting Leefbaarometer scores (Levering et al., 2023).

Furthermore, one can use contrastive methods, which compare timesteps or data points within each step with each other (Mazzaglia et al., 2021). Similarity loss tunes the neural network to capture how these data points are different. The space between the points, as it were. However, this requires a sampling heuristic, which may be difficult to get hold of in practice.

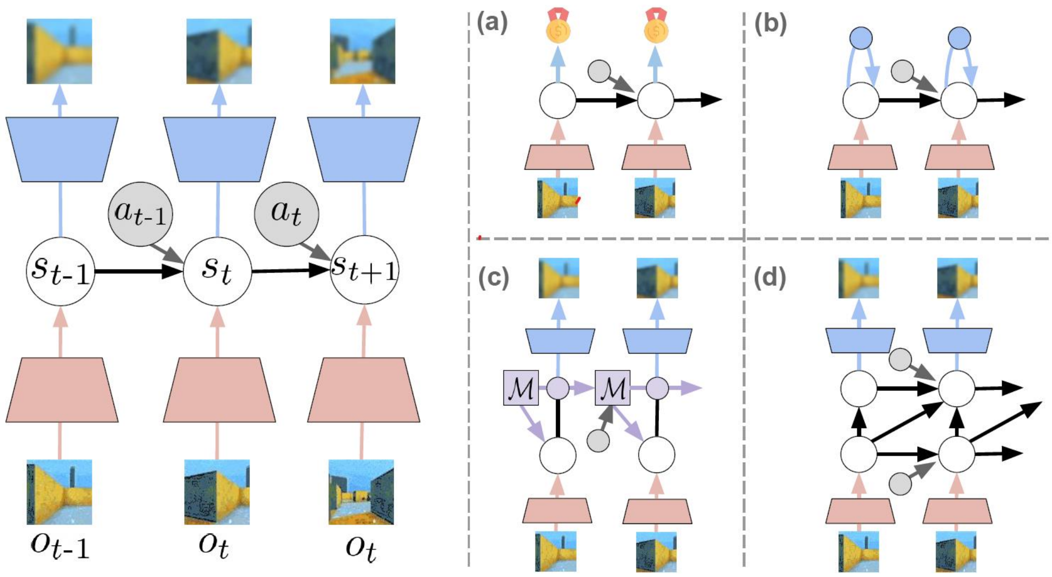


Figure 23: The free energy principle for perception and action: a deep learning perspective. Left: reconstruction of input data using an auto-encoder. Panel a: task-oriented representation. Panel b: state-consistent representations. Panel c: memory-equipped model. Panel d: hierarchical structure. From (Mazzaglia et al., 2022).

However, there is a trade-off between learning from raw data and using pre-processed metric representations. While raw data can capture more information, pre-processed representations are computationally efficient and valuable when the ambiguity of the mapping is low. Pre-learned metric representations (learnt by artificial neural networks) can simplify generative models by serving as prior knowledge, speeding up learning, and improving generalisation. This approach also supports transfer learning between different tasks or domains, making it a practical strategy in complex urban environments.

2.6 Conclusions Part One of Theoretical Framework

The first part of the theoretical framework covered the definition and operationalisation of liveability, revealing the concept's complexity due to its relation to terms such as well-being, environmental quality, quality of life, and sustainability. Two main approaches to understanding liveability have emerged: static and dynamic. Both approaches consider liveability as the fit between the resident and the living environment, but they differ in how they conceptualize and measure this fit.

The static approach, currently dominant in practice, relies on indicators and valuation to operationalise liveability. It takes a perspectival measurement of the resident-environment relationship, focusing either on the environment from the resident's perspective (liveability) or on the resident from the environment's perspective (quality of life). This approach has been formulated over the years to serve as a measurement instrument for policymakers, exemplified by operationalisations like the Leefbaarometer in the Netherlands.

In contrast, the dynamic approach, grounded in ecological psychology, views residents and the environment as a single, dynamically coupled system. This approach introduces concepts like niche construction, where residents actively shape their environment to construct suitable living spaces. In this view, accessibility is reframed as a landscape of affordances distributed across space and time, and social phenomena like social cohesion are seen as forms of cultural niche construction.

At this point, the first two research questions can be answered, and they aim to bridge the gap between the role of representations and liveability. The first question addresses the link between liveability and the action-perception loop, and the second question is the link between the action-perception loop and representations.

An answer to the first question is that the action-perception loop defines the dynamic approach. The dynamic approach involves actively palpating the outside environment with the expectation of certain observations. These are expressed and parameterised by probability distributions and paths of least action, solving Lagrangian equations. Furthermore, the fit between resident and their living environment is found between percepts of indicators and their needs/desires. Hence, perception is a proxy for fit, an ongoing dynamic process mediated by internally driven characteristic states.

On the other hand, the static approach to liveability places fit to occur between percepts of indicators and needs/desires. Fit is now a matter of valuation. Where perceived indicators, percepts, are weighted against their ability to satisfy needs/desires. There is no action-perception loop within the resident environment relationship.

An answer to the second question is that representations play an instrumental role in the action-perception loop. The dynamic approach to liveability is the action-perception loop, a process. Representations in the dynamic approach are outcomes that facilitate the process, not outcomes that stand by themselves. A generative model of the action-perception loop needs to have some parameters that encode the beliefs about internal and external states. The updates of those parameters are outcomes of dynamically selected actions and their forthcoming percepts. The static approach to liveability does not rely on the action-perception loop, and as such, any representation used to approximate perception is an outcome of a one-directional measurement. In turn, that percept is weighted off against needs/desires for valuation.

In line with the concept of *natura naturans* and *natura naturata*, the dynamic and static approaches are complementary, so they can be used to inform modelling decisions of each other. In this thesis, that means that the static approach, as operationalised using the Leefbaarometer, can be used to inform how to model the dynamic approach in the future. Since operationalising the dynamic approach as a whole is out of scope, the focus is on just the neural network, which maps high-dimensional indicators towards lower-dimensional percepts. Such a neural network is like a transmission to the information engine; the total assembly performs useful work by selecting, constructing, and maintaining its niche.

There are three key modelling considerations from the first part of the theoretical framework. 1) A higher spatial resolution is preferred since transportation impacts liveability strongly, and transport phenomena covariate across space such that more detail considerably improves performance (Miller et al., 2013b). 2) Various data sources have shown good performance in automating the creation of percepts using neural networks. Though mainly aerial- and street-view images perform the best. 3) Contrastive loss can be challenging to implement due to the need for heuristics. Luckily, transport planning has studied relevant heuristics for decades. Location-based accessibility aligns with affordances and the constrained maximum entropy principle (dual to the free energy principle), making it suited to use as a heuristic in contrastive sampling.

In sum, constructing the dynamic approach to liveability requires both a transmission and an engine, see Figure 24. Part one of the theoretical framework explains the engine while noting that representations are instrumental to its functioning (percepts). However, the acquisition of representations has not yet been addressed. Part two of the theoretical framework will, therefore, address the transmission of this broader modelling framework such that the methodology and results chapters can develop the transmission in further detail. The transmission involves extracting features from data which serve as indicators. Subsequently, these indicators, which themselves are representations, should be combined to align with maximum entropy, as the selection of actions from affordances applies constraints.

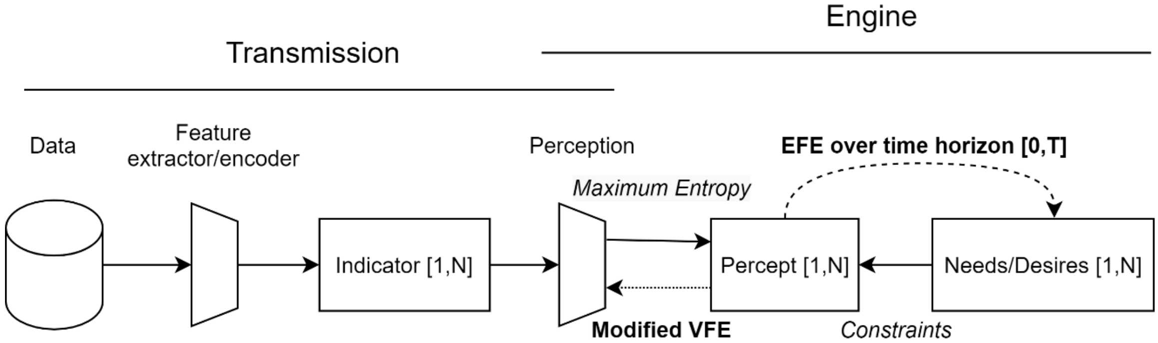


Figure 24: Information engine with transmission.Total assembly.

2.7 Representation Learning

The field of representation learning focuses on processing high-dimensional data into lower-dimensional representations. A specialized subfield, metric learning, introduces a metric distance by which objects are separated and distinguishable from each other in a vector space. Common metric distances include Euclidean, Manhattan, cosine angle, and dot product. For simplicity, we will refer to both representation learning and metric learning as representation learning and use the term "embeddings" for representations. Embeddings have three practical properties. 1) As vectors, they allow standard operations like addition, averaging, and multiplication. 2) Embeddings can be used in various machine learning settings, including non-deep learning algorithms. 3) Dimensionality reduction preserves the largest variation within the dataset, allowing for compact representations. There are linear and non-linear reduction techniques. Both have in common that it is possible to decompose high-dimensional data into fewer dimensions while retaining as much variation within the data as possible. For example, a linear technique, principle component analysis (PCA), learns a new coordinate system within the data, shifting and rotating the axis to give each data point a new coordinate.

The creation of embeddings relies on deep neural networks' ability to compress information. These networks map data to a latent vector space (manifold), representing data points with metric distances corresponding to their semantic similarity. Semantic similarity is context-dependent, measuring how close objects are to each other given all other objects in the dataset. The mapping function can be expressed as:

$$Embedding^r = Data^R * f(\varphi)$$

Where r is the lower dimension of the embedding, R is the higher dimension of the original data, and f(ϕ) is a neural network parameterised by network weights ϕ.

Neural networks used for this purpose are typically funnel-shaped (encoders), with fewer parameters at the output than the input. The embedding is extracted from the middle layer when data is fed through the network. Key considerations in developing deep neural networks for representation learning include:

- Data type (e.g., images, text)
- Learning strategy (self-supervised learning is common in metric learning)
- Neural network architecture
- Loss function (depends on the learning strategy)

This study focuses on self-supervised learning strategies, as they allow for learning representations useful for various downstream tasks without requiring target values to serve as ground truth. Recent findings suggest that different architectures using similar data tend to converge on similar representations (Huh et al., 2024). The loss functions used in self-supervised learning are typically either reconstruction-based (aiming to recreate the original input) or contrastive (sampling data points heuristically to learn similarities and differences). The upcoming section on urban representation learning will detail the specific types of data, neural network architectures, and loss functions used to represent urban regions. This two-stage process of learning representations and then applying them for verification is standard practice in metric learning studies (J. Wang & Biljecki, 2022).

2.8 Urban Representation Learning

In line with the focus on self-supervised learning strategies, this section focuses on urban representation learning studies that involve a two-step process. One of embedding creation and a subsequent task. Alternatively, if end-to-end learning is applied, the target data of the task of interest is included in the training of the neural networks from which embeddings are extracted.

Self-supervised urban representation learning draws upon two inductive priors. First, cities and their geospatially located data can be regionalised into smaller discrete spatial units, such as administrative units or evenly distributed geometric shapes like squares or hexagons. These spatial units provide the necessary classes and labels for self-supervised learning approaches. Second, the first law of geography states that things closer are more similar than things farther away (Tobler, 1970), guiding the training process to enforce semantic similarity of embeddings based on spatial proximity. It is essential to note that the first law of geography is to be interpreted loosely. The argumentation for this loose interpretation is that this inductive prior aims to ensure that the metric similarities of embeddings are aligned with geographic similarities of the urban region. The definition and operationalisation of a geographic similarity encompasses everything studied in geography and network science (graphs). Not just fly like a crow distance. Alternatively, it could be flows of, for example, vehicles, cargo, pedestrians, cyclists, ideas (Alex Pentland & Pentland, 2020), telecommunication data, water, pollutants, and cultures/narratives.

Three modelling decisions shape urban representation learning: the choice of spatial unit, data sources, and learning strategy. These decisions overlap significantly. For example, the chosen spatial unit constrains the set of applicable learning strategies, with uniformly tiled spatial units affording different methodologies compared to irregularly shaped ones.

The choice of discrete spatial units is primarily influenced by data availability. Many studies rely on administrative units, as most data is gathered at this level, ranging from countries to postal codes. Alternatively, geospatial operations like the Voronoi methodology are used in networked systems such as telecommunications (Almaatouq et al., 2016). Discrete global grid systems (DGGS) offer another approach (Kmoch, Matsibora, et al., 2022), using simple, tileable shapes to cover the Earth's surface. Each bit of global surface has a designated identifier, allowing for rapid calculations compared to conventional geospatial operations based on geometries. The H3 geospatial index developed by Uber has become a standard in geospatial machine learning, being taken up in several Python libraries, including SRAI (Gramacki et al., 2023). Its hexagonal format offers benefits over square indexing, including uniform distance between edge and centroid and better shape retention across latitudes (Kmoch, Vasilyev, et al., 2022).

Furthermore, H3 is hierarchical, in line with the nested nature of active inference models. Parent and child units are subsequently above or below a certain resolution. Hierarchical processing is highly promising as it aligns with the nature of geographic processes (Bejan & Lorente, 2010). Figure 26 provides an overview of four relevant regionalisation methodologies to create discrete spatial units, while Figure 25 illustrates different resolutions of the H3 geospatial index. Note the nested structure such that higher resolutions fit within lower ones.

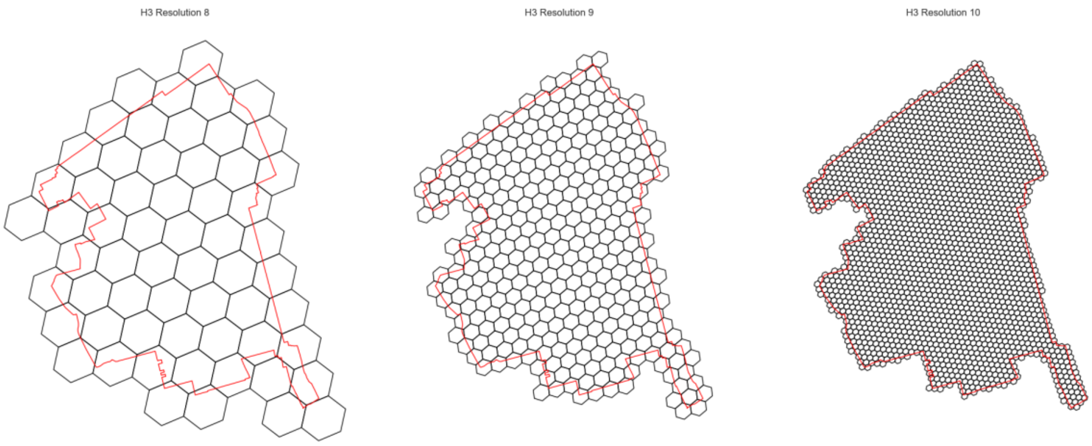


Figure 25: Visualisation of resolution in the H3 geospatial index system. Higher resolutions have a finer tiling.

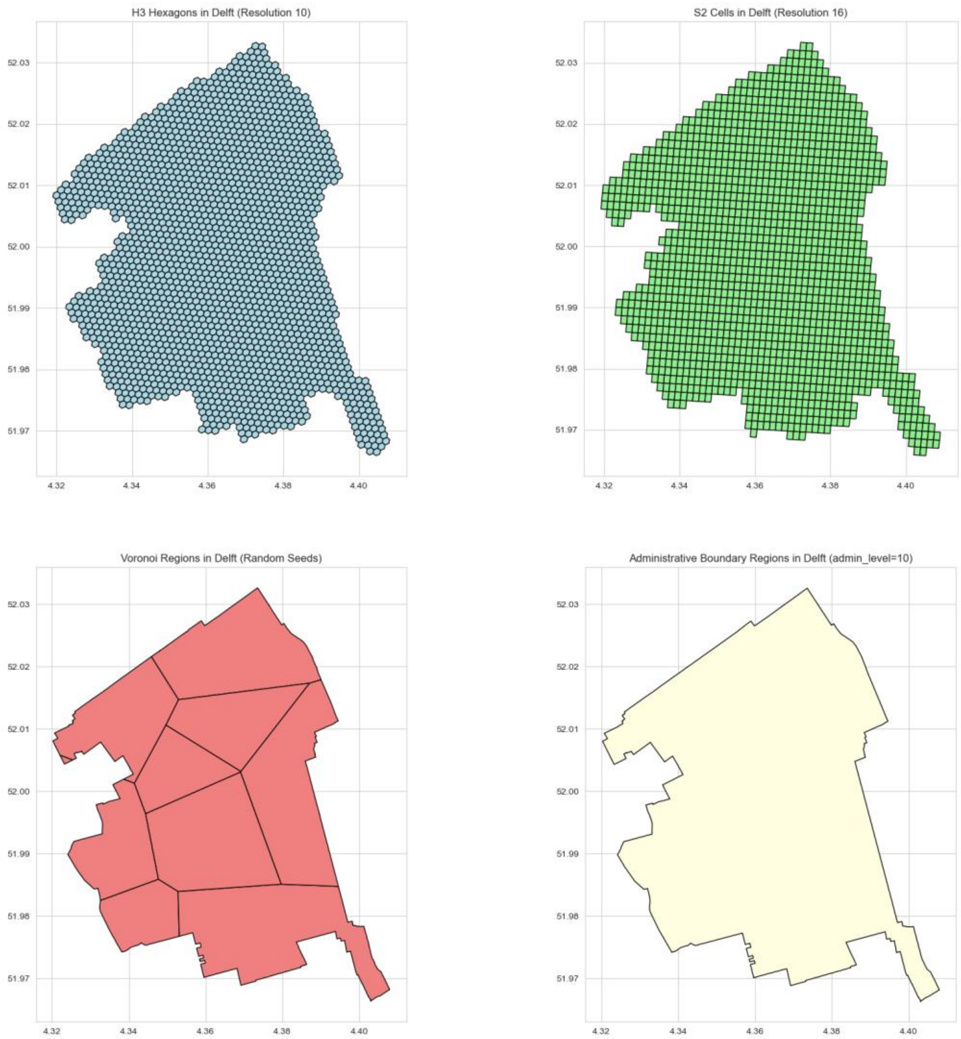


Figure 26: Four relevant regionalisation methodologies to create discrete spatial units. H3 hexagons, S2 squares, Voronoi, administrative borders.

Data sources in urban representation learning aim to approximate the urban region as a representational object in metric space. The data chosen should be varied over space sufficiently for the neural network to learn differentiating regions. Street view images have proven rich in information, correlating with various urban metrics (Fan et al., 2023; Huang et al., 2021; Z. Wang et al., 2020). Aerial and satellite imagery, while less common in urban representation learning (Jean et al., 2018), could draw on extensive experience from remote sensing. Points of interest (PoI) data and road network representations further enrich urban models, often synergizing with image-based data (Huang et al., 2021; Z. Wang et al., 2020). Some studies use hierarchical graphs for road networks (Hu et al., 2023; Wu et al., 2020), while others learn to assign features related to land use (Hu et al., 2021) or represent road types in spatial units (Leśniara & Szymański, 2022). Specialized data sources like building footprints (Li et al., 2023) and transit feed specifications (Gramacki, 2021) offer the potential for more nuanced representations.

The learning strategy encompasses both the loss function and neural network architecture. It involves embedding individual data sources and combining them. These operations may not always rely on neural networks, as elementary operations like sum, multiplication, and concatenation are often sufficient. Loss functions are typically reconstruction-based (e.g., mean squared error) or similarity-based, like triplet loss (Hoffer & Ailon, 2015) or circle loss (Y. Sun et al., 2020). Neural network architectures range from multi-layer perceptrons to convolutional networks and transformers, with some incorporating U-nets or auto-encoders. Graph neural networks are well-studied for operating on graph-structured data (Kipf & Welling, 2017). Applied to transport networks, these have shown promising results in inductively predicting the economic performance of metro stations' catchment areas (Xiao et al., 2021). The finding is that richer link information, such as passenger flows, performs better than distance information.

Addressing spatial context is consistent in urban representation learning, often referencing the first law of geography. Spatial context may be accounted for in two ways. First, sampling heuristics address spatial context through sampling similar and dissimilar regions. Second, neural networks with inductive bias, such as convolutional neural networks, have an inductive bias that is applicable to urban settings (Liang et al., 2022). The addition of graphs expands the range of options for modelling, particularly useful for transportation networks (Huang et al., 2021; Xiao et al., 2021). In terms of sampling, current research often uses Euclidean distance (Huang et al., 2021) or passenger volumes to sample similar and dissimilar spatial units (Huang et al., 2021; Luo et al., 2022; F. Sun et al., 2023; Xiao et al., 2021). Passenger volumes, often proxied by taxi trips, are generally viewed as better indicators of similarity than Euclidean distance. However, no study has explicitly considered accessibility as a measure of similarity in relation to the first law of geography.

2.9 Accessibility

Accessibility, previously discussed in relation to liveability, is now approached from a modelling perspective. It describes the distribution of activities across space and the travel resistances separating them. Within transportation policy, accessibility is a core topic as it relates directly to the transportation network's ability to meet residents' mobility needs.

In the field of urban representation learning, researchers use various measures to understand the similarity between different urban areas. Commonly, these include simple metrics like Euclidean distance or data on passenger flows. However, despite its importance in transportation policy, accessibility has not yet been used as a measure of similarity in urban representation learning.

Accessibility modelling involves several trade-offs and can be understood through various approaches. A fundamental distinction exists between aggregate and disaggregate approaches. Aggregate models consider zones within the transport network and align with macro-simulations of traffic networks. Disaggregate models, on the other hand, focus on individual travellers or population segments, often demarcated by socio-demographic characteristics such as gender or age. Another significant modelling decision is whether to employ decision theory, using a utility measure to quantify the willingness to use infrastructure as a measure of accessibility. While some methods do not explicitly use choice models, they may implicitly do so by estimating different parameters for segments of populations.

Geurs & van Wee (2004) define four types of accessibility. Infrastructure-based accessibility considers the characteristics of the transportation network, such as travel time (reliability), speeds, and congestion. While this approach provides a rich overview of network conditions, it does not specify how activities are distributed across time and space. Person-based accessibility, a disaggregate approach, evaluates travel patterns across time and space by modelling individual travel trajectories. This method is particularly useful for considering temporal edge cases, such as the often reduced public transport frequencies during off-peak hours. However, it faces practical challenges related to computational power and data availability.

The log sum approach, commonly used by transportation engineers, is particularly powerful (Hansen, 1972). It can calculate the consumer surplus of a transport policy intervention, making it suitable for monetary evaluation (van Wee, 2016). This approach is based on the concept of expected utility, assuming that people are (boundedly) rational choice-makers with varying preferences for different travel attributes. It involves summing all of the alternatives in the choice set, like travel modes or destinations. While it provides rich insights, the log sum approach requires extensive data on both travel resistances and traveller preferences. Furthermore, it may best be understood as the first three steps in the four-step transport model, including trip generation, trip distribution and modal selection. Only the fourth step, the assignment of flows across the network, is left out (Mink, 2023).

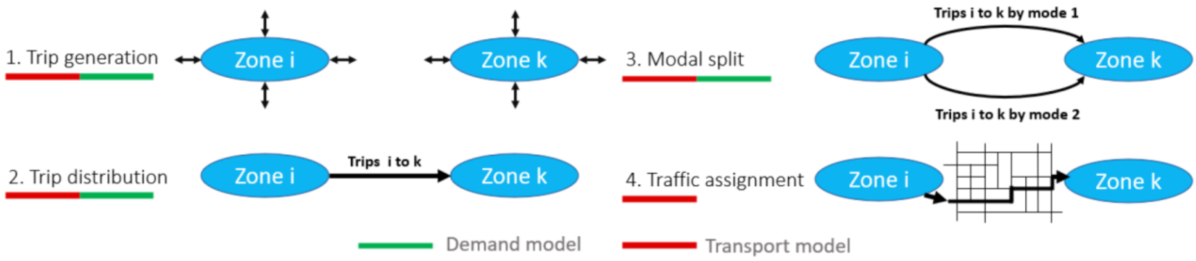


Figure 27: four-step transport model. Step 1: trip generation at every zone. Step 2: distribution from every zone to all others. Step 3: modal split decision of trip between zones. Step 4: assignment of trip potential throughout the transport network. From (Mink, 2023)

Location-based accessibility, favoured by geographers, captures the interaction between transport and activities across urban regions without considering individual traveller characteristics. It uses resistance curves to model how the likelihood of taking a trip diminishes with increasing travel costs. Geurs (2018) notes that power law functions often empirically fit these curves best while noting that exponential functions are more popular in practice due to theoretical roots in maximum entropy (Reggiani et al., 2011). This approach can be adapted for different population segments or travel modes and has been found to yield similar outcomes to the log sum approach when equally segmented. The definition of activities is a crucial modelling decision in accessibility studies, as they are the reason for travelling. Common approaches include counting jobs and shopping opportunities, inventorying desired services (K. Geurs & van Wee, 2023), or using building density as a proxy for activity intensity. For the latter, Harbers (2022) has developed automated indicators using ratios like the Floor Space Index (FSI) and Ground Space Index (GSI). Both consider parcel sizes in relation to the summed floor space across levels.

Interestingly, location-based accessibility has potential ties to active inference interpretations of perception. It aligns with the maximum entropy principle, which states that the most likely distribution of activities is the one with the least strong conviction of specific values. Wilson (1971) proposed the classic gravity model, which is derived from the entropy-maximising framework. While not apparent at first glance, the constrained maximum entropy and free energy principles explain the same phenomena (M. J. D. Ramstead et al., 2023)— maximising entropy given constraints or minimising free energy given a generative model. A second duality is the perspective shift between internal and external states. Most notably, this duality means that self-organisation (life) occurs because the world is dissipative, as increased order in small pockets allows for more disorder everywhere else.

In the context of perception as a dynamic process, bottom-up errors force an open mind while top-down priors apply constraints. The interaction between these processes forms the act of perception. When modelling urban regions, using location-based accessibility to guide the self-supervised learning of representations may offer valuable insights. To illustrate, we can simulate the location-based accessibility of schools in Delft, Figure 29. When constrained to prefer only two randomly sampled schools, Figure 30, we see a completely different landscape of affordances compared to an unconstrained scenario, Figure 31. It demonstrates that while accessibility provides the affordance for travelling, the resulting landscape is subjective, shaped by prior preferences. In essence, accessibility offers a way to model not just the objective structure of infrastructure but also how it is perceived and used by residents with different preferences. This view is aligned with the capabilities approach in broad prosperity.

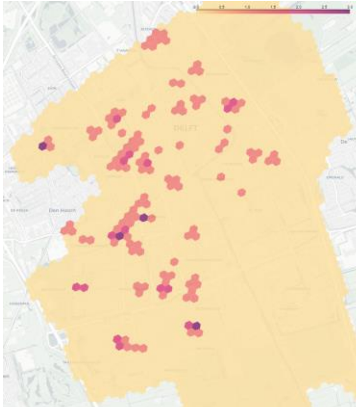


Figure 28: Number of schools per spatial unit for Delft.

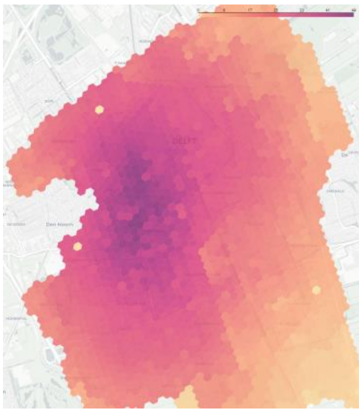


Figure 29: Location-based accessibility for schools using walking infrastructure in Delft.

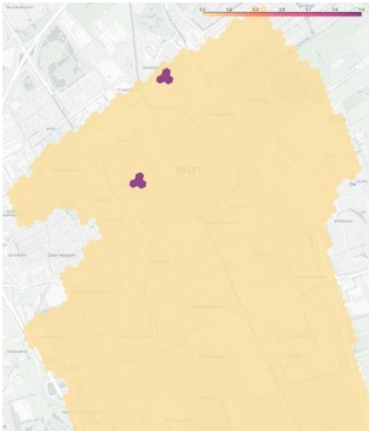


Figure 30: The constrained set of schools per spatial unit in Delft.

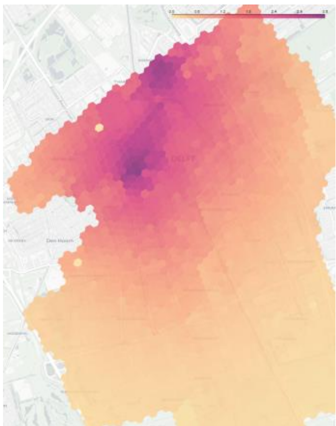


Figure 31: Location-based accessibility for a constrained set of schools using walking infrastructure in Delft.

2.10 Encoding views of urban data

In urban representation learning, individual data types are often referred to as 'views', each representing a distinct perspective of the urban system. Combining different views is known as multi-view learning. Recent open-source efforts, such as those by (Gramacki et al., 2023), aim to provide comprehensive toolboxes for creating representations of these individual views. The process of encoding individual views follows a four-step procedure: First, data is downloaded and stored with geospatial coordinates. Second, this data is assigned to corresponding spatial units based on spatial overlap. Third, an encoder neural network is trained on the discretized data. Finally, the data is fed through the encoder network again to extract representations.

The representation of POI data has its roots in word embedding techniques, particularly word2vec (Mikolov et al., 2013). This approach creates a dictionary where each word—POI—is assigned an embedding, with metric representations indicating similarity to others. Word2vec uses two main approaches to account for context: skip-gram and continuous bag of words (CBOW). Recent alternatives include GloVe (Pennington et al., 2014) and larger language transformer models. Building on word2vec, Woźniak and Szymański (2021) developed hex2vec, a convolutional-like approach using H3 hexagons. Donghi and Morvan (2023) further improved this with Geovex, which uses a convolutional auto-encoder and incorporates hexagonal convolutions and a Poisson distribution to learn the presence and absence of POI labels. Some studies, like Hexaconv, have also developed convolutional neural networks for hexagonal gridded data (Hoogeboom et al., 2018; Zhao et al., 2021).

Image encoders can be broadly categorized into two groups: those creating embeddings and those assigning classes of objects through segmentation. The former often relies on established convolutional neural network architectures like ResNet and AlexNet, typically trained on large datasets like ImageNet (He et al., 2015; Krizhevsky et al., 2017; Russakovsky et al., 2015). While these networks aim to classify objects, it is possible to extract metric representations from their penultimate layers. Image segmentation approaches, on the other hand, preprocess images into more interpretable formats. (Fan et al., 2023) used this method to represent urban regions by counting objects in different classes, while (Gong et al., 2018) identified 'street canyons' by mapping the presence of sky, trees, and buildings.

The SRAI Python package (Gramacki et al., 2023) incorporates several specialized encoders for H3 hexagons. One such encoder deals with the generalized transit feed specification, using an auto-encoder to compress various attributes of public transport timetables (Gramacki et al., 2021). By focusing on stops assignable to hexagons, the analysis captures the total number of stops made within that hexagon (summed across stops if multiple are present), number of directions (routes), and frequency of vehicles leaving per hour (6:00-22:00). Another, highway2vec, encodes road network characteristics, accounting for network links that may intersect multiple spatial units (Leśniara & Szymański, 2022).

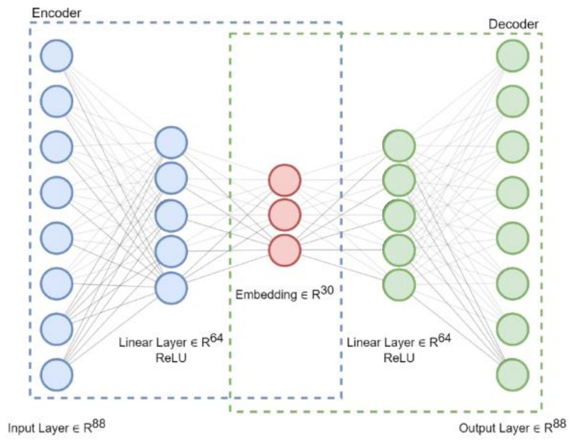


Figure 32: Schematic of an auto-encoder emphasising the embedding in the middle. From (Leśniara & Szymański, 2022).

Graph structures, which explicitly contain relationality, can be encoded using either shallow or deep approaches. Shallow encoders like DeepWalk (Perozzi et al., 2014) and node2vec (Grover & Leskovec, 2016) use random walks to contextualise neighbourhoods. While computationally expensive for large graphs, they are straightforward and can capture structural similarity. Deep graph encoders share aggregation weights across local neighbourhoods, employing techniques like convolutions (Kipf & Welling, 2017) or attention mechanisms (Veličković et al., 2018). Recent research has focused on improving the efficiency of these encoders to handle million-node graphs, with examples like GraphSage applying sampling techniques (Hamilton et al., 2018).

2.11 Learning Across Multiple Views

Learning representations from multiple data modalities, or views, is a significant obstacle in urban representation learning and deep learning in general. Multi-modality offers several advantages: representations become richer by combining correlations from different sources, each modality captures unique distributions, and the impact of noise may be reduced as correlations between modalities are preserved. However, these benefits have drawbacks, primarily the difficulty in engineering techniques to fuse views effectively and the need for larger, more resource-intensive models. At the heart of this challenge lies a broader technical discussion in metric representation learning: should encoder networks be fine-tuned to task-specific objectives? This debate has led to the development of various strategies, broadly categorised into data-centric fusion and learnt fusion approaches.

The data-centric fusion approach, exemplified by relative representations, offers a computationally efficient alternative to fine-tuning. Norelli et al. (2023) argue for leaving encoders to extract features without fine-tuning, proposing a data-centric method to align outputs of frozen pre-trained multi-modal encoder networks. Relative representations, introduced by (Moschella et al., 2023), embody this philosophy. They were developed as an alternative to resource-intensive models like CLIP (Contrastive Language-Image Pretraining), which require industrial-scale computing and data to train (Radford et al., 2021). Relative representations work by leveraging the fact that different modalities often describe similar classes of objects. They require only a set of absolute representations (raw output from an encoder network) and class labels. A similarity metric is then used to sample and compare representations across classes, resulting in a scaled set of vectors for all data points. This method can even combine encoder/decoder networks of different architectures. However, relative representations have limitations: they depend heavily on the encoder networks' training distribution, lose distance information between points, and can become computationally expensive for large datasets with many classes. Preliminary experimentation in the current study found that different data sources do not overlap enough in urban settings for relative representations to work.

On the other hand, learned fusion approaches involve training an additional network to combine representations from different views and spatial units. While simple vector operations like addition and averaging can be used for fusion, as demonstrated by (Raczycki, 2021) with hexagonal spatial units and concentric rings, most urban representation learning methods implement some form of neural network for fusion (Chan & Ren, 2023; Kim & Yoon, 2022, 2022; Li et al., 2023; Liang et al., 2022; Luo et al., 2022; Xiang, 2020). The main reason is that simple vector operations cannot account for correlations across views within individual regions (F. Sun et al., 2023). Instead, the references above propose various approaches to fuse multi-view data using attention networks (Vaswani et al., 2017). Attention is commonly used as it is highly versatile and flexible in capturing correlations, yet it is computationally expensive and needs a large dataset to achieve good training results. Graphs are readily employed in fusion network architectures for urban data as they are agnostic to the type of spatial unit. However, this may also make these studies harder to interpret due to the reliance on mathematical formalisms over intuitive architectures. Many of these methods employ multi-stage fusion, where within-region correlations across views are accounted for before addressing between-region correlations (Chan & Ren, 2023; Kim & Yoon, 2022; Li et al., 2023; F. Sun et al., 2023). The temporal component may also be represented, such that passenger flows used to steer the self-supervised learning approach account, e.g. morning and evening commute (Kim & Yoon, 2022).

All multi-view fusion methods in the urban representation learning literature employ a reconstruction loss to compress the high-dimensional data. However, recent work shows that reconstruction and perception differ, indicating that representations learned using reconstruction loss are unsuited for perception (Balestriero & LeCun, 2024). That is, features learnt using reconstruction loss while explaining much variance in the data perform poorly on image recognition. Moreover, liveability requires perception rather than pure compression persé; after all, one is perceiving indicators with needs & desires in mind.

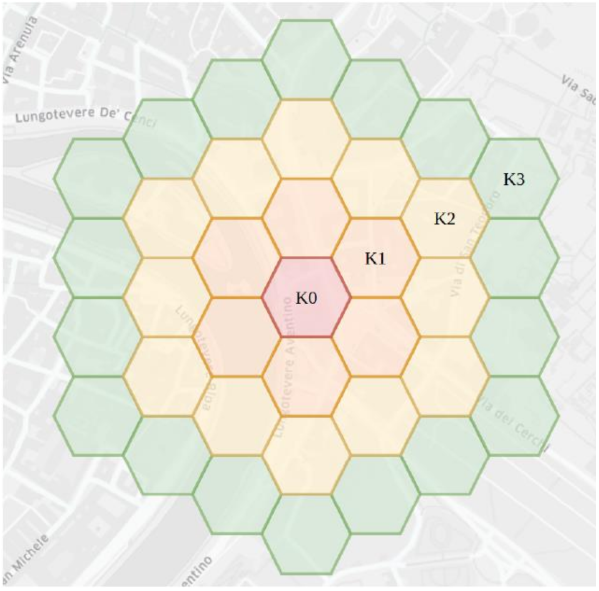


Figure 33: Illustration of concentric k-rings around a central H3 hexagon. From (Raczycki, 2021).

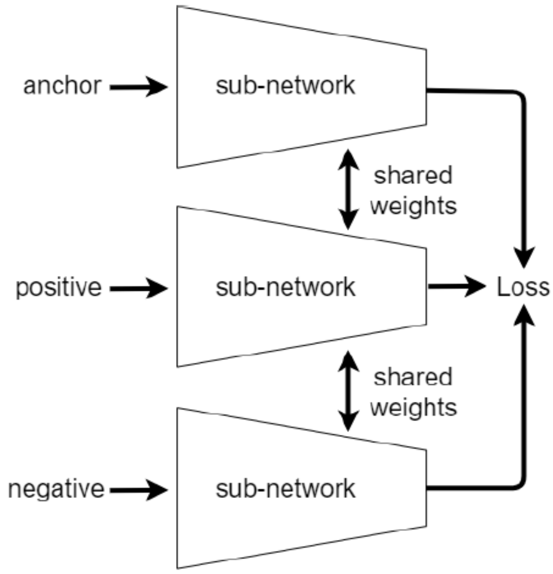


Figure 34: Triplet networks for contrastive loss. From (Ghojogh et al., 2022a).

An alternative to reconstruction loss is those based on similarity, such as triplet (Hoffer & Ailon, 2015) or circle loss (Y. Sun et al., 2020). Urban representation studies applying similarity loss opt for fine-tuning approaches, sequentially adding information from different views through loss functions (Huang et al., 2021; Z. Wang et al., 2020). The basis of similarity learning is the reliance on triplets, which is done heuristically and hence up to the modeller (Garrido et al., 2024). Unfortunately, little consideration is given to these heuristics in the urban representation literature. Commonly used measures are Euclidian distance and passenger flows (taxi trips). In the sequential similarity loss approach, encoder networks are tuned to consider correlations with other data modalities, resulting in features that account for multiple modalities. The resultant urban embedding of a spatial unit from a previous step is used in subsequent ones as initialisation. Calculating similarity loss is done by gathering triplets of data points, sampled using heuristics. The difference in metric distance (not spatial distance) is fed into the loss calculation such that the positive distance is minimised and the negative distance maximised. Circle loss improves on triplet loss by weighting these two metric distances differently depending on their relative size.

$$Triplet\ Loss(a, p, n) = max(0, \Delta\_p) - \Delta\_n + margin)$$

$$Circle\ Loss(a, p, n) = log(1 + sum(exp(\gamma(\alpha\_p - \Delta\_p))) + sum(exp(\gamma(\Delta\_n - \alpha\_n))))$$

where γ is a scale factor, α_p and α_n are to be calculated parameters for positive and negative pairs, and Δ_p and Δ_n are the similarity scores for positive and negative pairs, respectively.

2.12 Conclusions Part Two of Literature Review

The second part of the literature review provided an overview of metric learning prerequisites, leading to an outline of multi-view urban representation learning. Several modelling steps were identified, including selecting a spatial unit type, data sources, and learning strategy. Multiple approaches for learning representations across views were examined, ranging from simple vector operations to learning fusion networks or fine-tuning single-view encoder networks with information from other views.

It was determined that reconstruction loss should be avoided for urban representations aimed at perception, as required for liveability assessments. Conceptual links between accessibility and the free energy principle were observed, supporting a dynamic approach to liveability from a transport modelling perspective—active inference, a description of the dynamics as seen in the free energy principle, might be like a transport model based on accessibility.

A comprehensive list of modelling choices that underpin this thesis's methodology is presented below. This list aims to provide the ingredients needed to design the transmission for future work, incorporating it into an engine. Like the first three steps in the 4-step transport model, in which assignment is left for the dynamics of active inference.

Hexagons are the most suitable spatial unit due to their isotropy (almost round shape), improving the validity of accessibility measures (catchment area). Furthermore, the H3 geospatial index is hierarchical, making it future-proof for active inference modelling.

Spatial convolutions incorporate the first law of geography, which states that similar entities are closer together. The Leefbaarometer relies on different spatial operations involving weighted functions. Spatial convolutions can accommodate weighted averages when adjusted.

Location-based accessibility aligns with the constrained maximum entropy principle, making it suitable for models built using the free energy principle. Additionally, as found in conclusion part one, location-based accessibility aligns with the notion of affordances as studied in active inference and the free energy principle.

Of the available self-supervised training losses used in urban representation learning, circle loss was deemed preferable. Triplet loss is less likely to converge than circle loss. Reconstruction loss was found unsuitable for perception tasks.

Learnt fusion networks are preferred over sequential training of lookup tables. Late fusion is particularly applicable for future dynamic operationalisations of liveability. Late fusion involves encoding individual views of urban data before fusing them. Early fusion, on the other hand, immediately combines raw data. Two reasons underpin this finding. First, deep learning in combination with generative models is only practised with individual (fusion) networks, not sequentially. Second, late fusion is justified given the role of perception; mapping indicators to percepts in generative modelling must not be included but may be outsourced to neural networks if the ambiguity of the mapping between indicators and percepts is low. <u>As such, indicators can be representations.</u> Ambiguity relates to the likelihood mapping between indicators and percepts; mapping can be one-to-one or many-to-many. High ambiguity in the to-be-modelled phenomenon necessitates active data sampling to maximise information gain as found in the decomposition of expected free energy.

The road network, points of interest, street view images, aerial images, and public transport data are suitable data sources for representations of urban regions.

3 Methods

This chapter outlines the developed methodology to create urban representations using metric representation learning. We present a novel learning strategy for urban representations based on spatial convolutions and implement the work of other authors for comparison. The comparative works are known in the literature as Urban2Vec and M3G (Huang et al., 2021; Z. Wang et al., 2020).

The methodology is structured to start with an overview of the study area and its regionalisation, the data collection and preparation, the two learning strategies, and the setup of the experiments conducted.

The two conclusions of the theoretical framework contain choices underpinning learning strategy one. Each of the findings is incorporated into the preparation of the study area, the choice of data sources, and the entire development of the learning strategy.

Two study areas and their associated regionalisation are shared universally within this thesis, consistent for both learning strategies. However, a slight difference is that learning strategy one draws upon the buffered set of spatial units since it involves spatial convolutions requiring padding.

This study chooses the hexagon (H3) since it is high in isotropy, hierarchically structured, and has many established encoder models packaged in the SRAI Python library (Gramacki et al., 2023). Both learning strategies use a subset of spatial units selected for the presence of Leefbaarometer scores. The Leefbaarometer uses a square geospatial indexing system. Hence, direct intersection gives skewed results. A spatial unit is selected based on the surface overlap, conditional on this overlap being at least a hectare to preserve the 100x100m resolution of Leefbaarometer scores.

Additionally, the values of the Leefbaarometer are copied onto the hexagonal spatial units using a weighted sum based on the overlap on the surface. There is a significant difference in the granularity at different H3 resolutions. Where a variation in Leefbaarometer scores at resolution 10, on the left Figure 35, is much more pronounced than that at resolution 9, on the right Figure 36. Warmer colours have higher scores than cooler colours, which have lower scores.

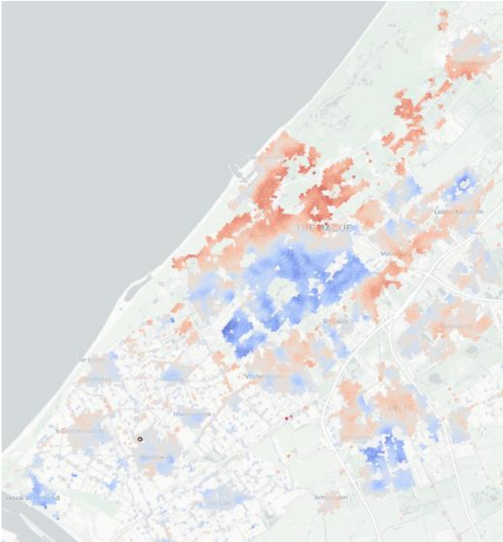


Figure 35: Aggregate Leefbaarometer score spatially joined with H3 resolution 10.

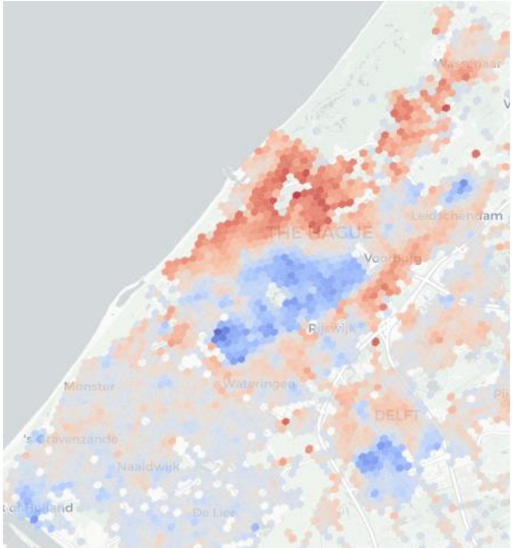


Figure 36: Aggregate Leefbaarometer score spatially joined with H3 resolution 9.

Additionally, learning strategy one uses spatial convolutions and requires additional padding of spatial units. A buffered set of spatial units is created and to be assigned data, as will be elaborated in section two. The buffer is five rings of hexagons for H3 resolution 9 and fifteen for resolution 10. See Figure 37 and Figure 38.

Figure 37: Regionalisation at H3 resolution 9 for spatial units with Leefbaarometer scores.

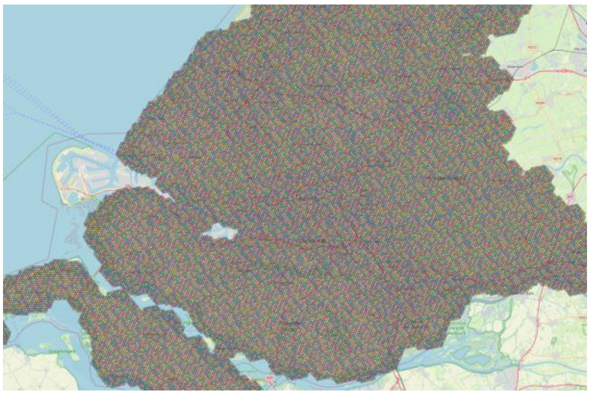


Figure 38: Regionalisation at H3 resolution 9 for spatial units buffering residential areas.

3.1 Data Collection

Data collection can be differentiated into two classes: intra-unit and inter-unit. Intra-unit data relates to data points with spatial locations, such as images or points of interest. Inter-unit data relates to the relationship between spatial units, such as distance or accessibility.

3.1.1 Intra-unit

The intra-unit process requires the collection of data points with their associated geospatial location for indexing into the corresponding spatial unit based on their coordinates. All data points have coordinates and spatial units cover a surface of coordinates. It is then a matter of assigning data points to their spatial unit based on their overlapping coordinates. For example, if a point of interest is within the polygon of a spatial unit, it is assigned to the spatial unit. In the case of linear geometries like roads or rivers, an intersection is used where it is sufficient for any linear geometry to intersect with a polygon of the spatial unit. If a line intersects with multiple spatial units, that data point's characteristics are assigned to all spatial units with which it intersects.

The different data sources—views—are the road network, general transit feed specification, points of interest, street view and aerial images. The road network is obtained from OpenStreetMap and relates to the car network. The general transit feed specification can be retrieved from its website, gtfs.org. Points of interest are obtained from OpenStreetMap with the Geofabrik downloader and its associated filter; for more details, see geofabrik.de. Street view images were retrieved from Google Street View using the methodology from (Garrido-Valenzuela et al., 2023). Valenzuela et al. assign panoids, 360-degree panorama views, of images to each H3 spatial unit, which are split into four sections of 90 degrees using the driving direction as reference such that the centre of each image is either the front, back or side of the vehicle. Aerial images are obtained from pdok.nl, an open dataset provided by the Dutch government. Publieke Dienstverlening Op de Kaart (PDOK) provides an Application Programming Interface (API), which can be called in a Python script. Furthermore, since images must fit into an image encoder, the size of these images is 224 by 224 pixels. As such, the hexagonal spatial units were given a square bounding box whose' coordinates, along with the image resolution, were queried to the API.

3.1.2 Inter-unit

Inter-unit data covers relationships between spatial units, like Euclidean distance and transport network resistances. We incorporate these resistances through location-based accessibility, combining travel times and building density as a stand-in for destination attractiveness.

To calculate Euclidean distance, one considers the distance between centroids of spatial units. Since these distances are relatively short, there is no need for distance measures that account for the earth's curvature, such as Haversine.

Calculating location-based accessibility requires travel times between spatial units and each unit's attractiveness. Whereas the number of jobs or shops is often used to quantify the attractiveness of spatial units, building densities are a suitable alternative. The RUDIFUN dataset contains indicators of building densities for the entire Netherlands. From these indicators, the floor space index (FSI) is well suited as it captures the floor space ratio to building footprint, including parking lots and other land use on the parcel. FSI scores are assigned to hexagons using a weighted sum based on the spatial overlap, just like applied when intersecting the Leefbaarometer scores in the preparation of the study area (Figure 40).

The location-based accessibility of a spatial unit is calculated by summing the factors of attractiveness and travel resistance of all other spatial units. Each spatial unit is once treated as origin (subscript i), at which point all other units are destinations (subscript j). Travel resistance is calculated using an impedance function that relies on a distance decay rate to penalise longer travel times, reducing the likelihood of trips between distant spatial units. See Figure 39.

$$A_i = \sum_j FSI_j * f(T_{ij})$$

In this study, we use an exponential decay rate applied to travel time in seconds with a cutoff of one hour (3600 seconds). The decay rate was made up in this study, though it is possible to use empirical values segmented for segments of a population for maximum accuracy.

$$f(T_{ij}) = e^{-\beta * Tij} = e^{-0.001 * Tij}$$

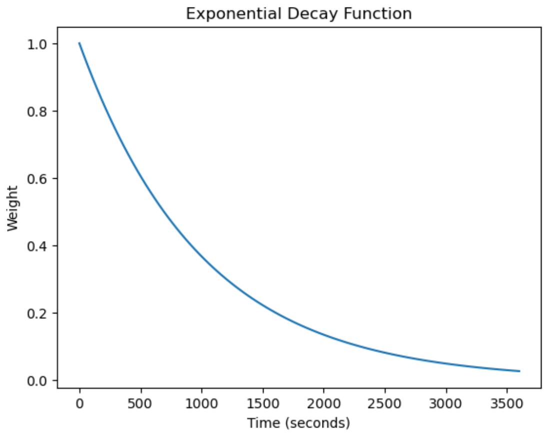


Figure 39: Distance decay function for travel resistances to be used in location-based accessibility.

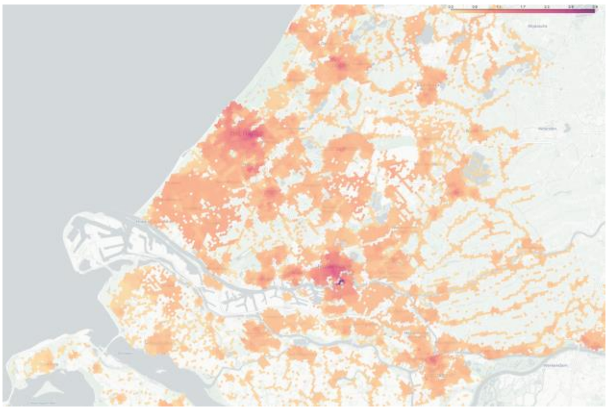


Figure 40: Plot of building density operationalised using floor space index (FSI).

In sum, location-based accessibility and Euclidean distance are calculated similarly, sharing the first step. After this, Euclidean distance is a simple calculation between centroids. Moreover, location-based accessibility continues with the remainder of the steps as outlined below.

1. Each spatial unit is the catchment area of its centroid.
2. For each travel mode, walking, cycling, and driving:
    a. Using OpenStreetMap networks for walking, cycling, and driving, network nodes are assigned to hexagon centroids based on closest coordinates. A K-D Tree is used to partition space into a searchable tree, improving computational efficiency.
    b. A single-source Dijkstra algorithm is used in a loop over all regions to create an origin-destination matrix of travel times. Compared to a doubly nested-for-loop, single-source dijkstra builds a reusable computation graph for all destinations, saving hours.
    c. Calculate location-based accessibility by applying the impedance function with decay rate and attractiveness.
3. Combine accessibilities for walking, cycling, and driving with equal weight, taking their average.

3.2 Learning Strategies

This thesis explores two learning strategies for creating urban representations. The former is developed in this thesis, whereas the latter is an adapted version of earlier work.

The first strategy, which we propose as a novel approach, relies on spatial convolutions. For each spatial unit, spatial convolutions aggregate information about their neighbourhood with the optionality to apply (learnable) transformations within the aggregation process. Spatial convolutions are like conventional geographic information system operations in that a feature is calculated given a distance towards other features. For example, the buffer zone around a transport corridor moving hazardous materials. In particular, the developed methodology draws inspiration from graph convolutional neural networks (no sliding window) but eliminates the need for explicit graph construction. Instead, it utilises the neighbourhood indexing provided by the H3 geospatial index, offering a more streamlined approach to capturing spatial relationships. The commonality here is that graph convolutional neural networks calculate the features of a central node using increments of hops throughout the graph. The features of each adjacent ring of nodes are transformed using a neural network and subsequently aggregated using an average or maximum thereof. Unlike graphs, hexagonal spatial units do not rely on incremental hops but can directly query rings around a centre since all spatial unit and their relationships are indexed. Hence, the name ring aggregation.

The second strategy is adapted from existing work, specifically Urban2vec and M3G (Z. Wang et al., 2020; Huang et al., 2021). Minor adjustments to the work of Huang and Wang et al. are necessary due to the difference in study area, altered choice of spatial units and data limitations. While it is impractical to copy the approach, capturing the generalities for comparison is feasible. Adjustments involve exchanging the San Francisco Bay Area for the province of South Holland, replacing neighbourhoods structured as nodes in a graph with H3 hexagons for the spatial unit, aerial images instead of street view images, and a different dictionary of point of interest labels used. These adjustments ensure that both learning strategies depend on the same H3 regionalisation and data for comparison.

3.2.1 Strategy one – Ring Aggregation

The first learning strategy developed in this thesis builds upon the ring aggregation methodology introduced by Raczycki (2021). Ring aggregation is like spatial convolutions due to the isotropic nature of the H3 geospatial index. For any randomly picked hexagon, the rings around it are uniform in distance and number (Figure 42), unlike graphs, which rely on adjacency matrixes. The isotropy allows for innovation in spatial convolutions by combining the work of Raczycki and graph convolutional neural networks. Whereas convolutions work by recursively applying transformation and aggregation per ring, working from the furthest ring towards the centre—ring aggregation calculates a transformed aggregation per ring and subsequently applies a secondary transformed aggregation across those rings.

This innovation allows for weighted averages across the rings and transformations operationalised using learnable neural network weights. At its core, this learning strategy is a late-fusion approach. It combines multiple views' representations so that each spatial unit is left with a single final representation. The fusion relies on a two-step process: first, take the average embedding per ring. Second, a weighted average across rings is applied to get the final central representation.

The mathematical formulation of our ring aggregation strategy is as follows: denote the concatenated representation per hexagon as Ri, the mean of a ring as Mk, its weight as Wk, and the resultant central aggregated embedding as S (averaged over K rings), then the central aggregated embedding is expressed as a sum over transformed weighted average of transformed ring means across rings:

$$S = \sum_{k=0}^{K} W_k * f_\theta(M_k)$$

The average per ring equals the inverse of the number of hexagons in that ring and the sum over transformed concatenated representations of spatial units within that ring:

$$M_k = \frac{1}{I} * \sum_{i=1}^{I} f_\varphi(R_i)$$

Combined, this gives an expression with two learnable neural networks. Both networks map one representation to another, reducing the dimensionality at each step. The networks are parameterised by weights theta for across rings and phi for within:

$$S = \sum_{k=0}^{K} W_k * f_\theta\left(\frac{\sum_{i=1}^{I} f_\varphi(R_i)}{I}\right)$$

The contribution across rings is defined according to four different weighting schemes: natural exponent, logarithm, linear, and flat:

$$W_k = e^{-k}, \quad W_k = 1/\log_2(k+2), \quad W_k = 1 - \frac{1}{K}, \quad W_k = \frac{1}{K}$$

While Raczycki (2021) employed the exponent, linear, and flat weightings, this study additionally includes a logarithmic weighted average. This addition aligns with the calculation of the physical living environment as operationalised by the Leefbaarometer.

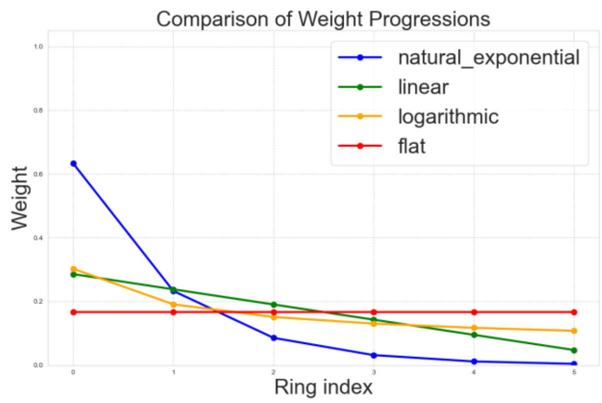


Figure 41: Weighted average types impact of each k-ring up until five rings. Blue exponential, green linear, orange logarithm, red flat.

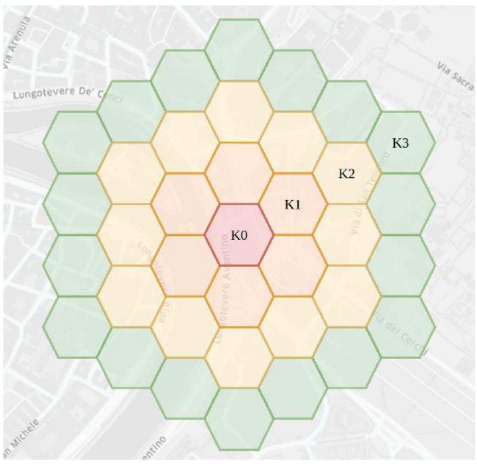


Figure 42: Illustration of concentric k-rings around a central H3 hexagon. From (Raczycki, 2021).

During preliminary experimentation, several observations were made. The dimensionality of hidden layers relative to input and output and the number of layers within each step had a limited impact on the model's performance. However, normalising the input data proved to be crucial. We also observed that using ReLU as an activation function in intermediate layers led to excessive sparsity.

A batch normalisation layer is added to address these issues, efficiently normalising data across modalities. This batch normalisation applies to the first layer, thus normalising the concatenated views per batch. Acceptable performance is achieved using a batch size of 256 and a learning rate of 0.0001 using the Adam Optimiser, which has no learning rate scheduler. Furthermore, ReLU was replaced with GeLU (Rectified error Linear Unit/Gaussian Error Linear Unit), as it prevents excessive sparsity of the representations, which impacts downstream performance on predicting the Leefbaarometer score negatively.

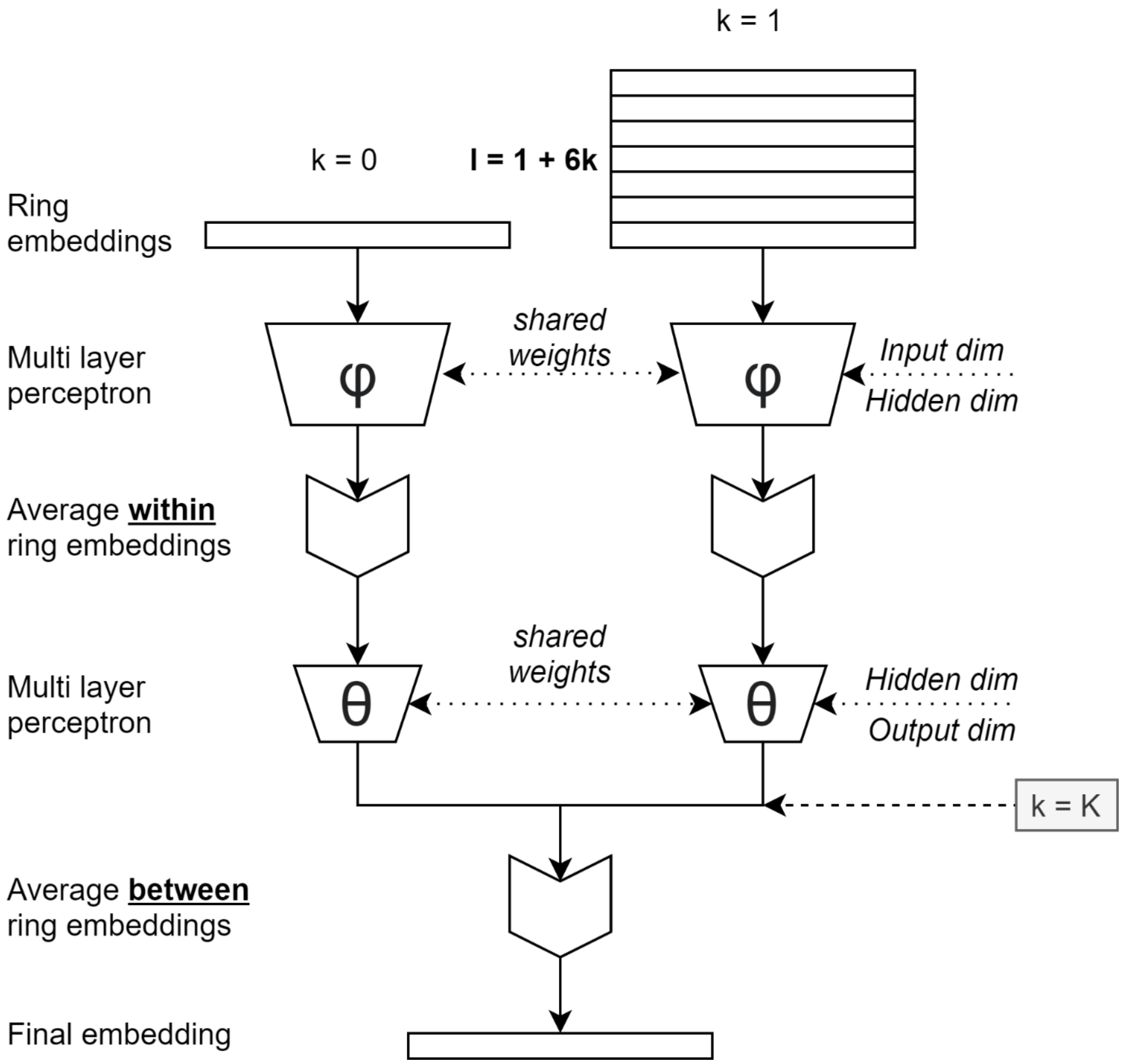


Figure 43: Learnable ring aggregation fusion network. Note that the number of input embeddings increases linearly with additional rings by a factor of six. Input dimensionality is reduced in two steps into a considerably smaller output dimensionality. Weights are shared for each step, denoted by theta and phi.

Next, a variety of encoders create embeddings. All encoders are trained or finetuned on their respective data covering the buffered study area; see Figure 38. Only street view images do not enjoy finetuning; instead, they use the pre-trained version of ConvNeXt based on ImageNet, while aerial images will be tested with and without a finetune—see next section on learning strategy two for details. Before being concatenated, the image embeddings are reduced in dimensionality to 100 to ensure roughly equal size for all embeddings.

Table 2: Encoder networks used to create indicators.

Data	Encoder Network	Training Data
Public transport stop characteristics	GTFS2vec (Gramacki et al., 2021)	Buffered study area
Road network characteristics	Highway2vec (Leśniara & Szymański, 2022)	Buffered study area
Aerial images	ConvNeXt Large (Liu et al., 2022)	Imagenet pre-train w/ optional finetune on buffered study area
Street-view images	ConvNeXt Large	Imagenet pre-train
Points of interest	Geovex (Donghi & Morvan, 2023) Hex2vec (Woźniak & Szymański, 2021)	Buffered study area

All embeddings are then concatenated (placed next to each other) and fed into the learnable fusion network. The learnable fusion network is trained using circle loss, with a gamma of 250 and m of 0.15. The triplets for training are sampled based on a measure of proximity. Given any anchor neighbourhood, a positive neighbourhood is in the top 2% - either lower than 2% of all pairs for Euclidean distance or better than 98% for location-based accessibility. A triplet contains three items, each fully aggregating the local spatial context. When sampling over five rings at resolution 9, there are 93 concatenated embeddings as input:

$$N\_hexagons_{triplet} = 3 * (1 + 6K) = 3 * (1 + 6 * 5) = 93$$

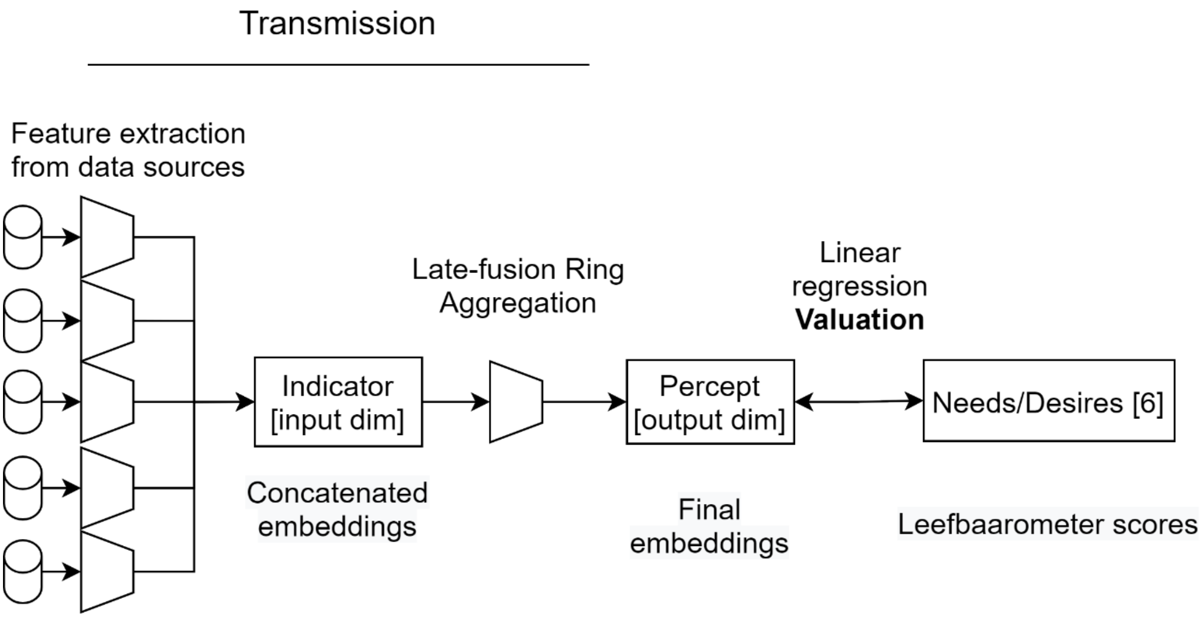


Figure 44: Static approach to liveability with indicators, percepts and needs/desires. Learning strategy one applies late fusion, such that concatenated embeddings of single views are the indicators. The percepts are a transformed aggregation of indicators. See experiments for linear regression with Leefbaarometer scores.

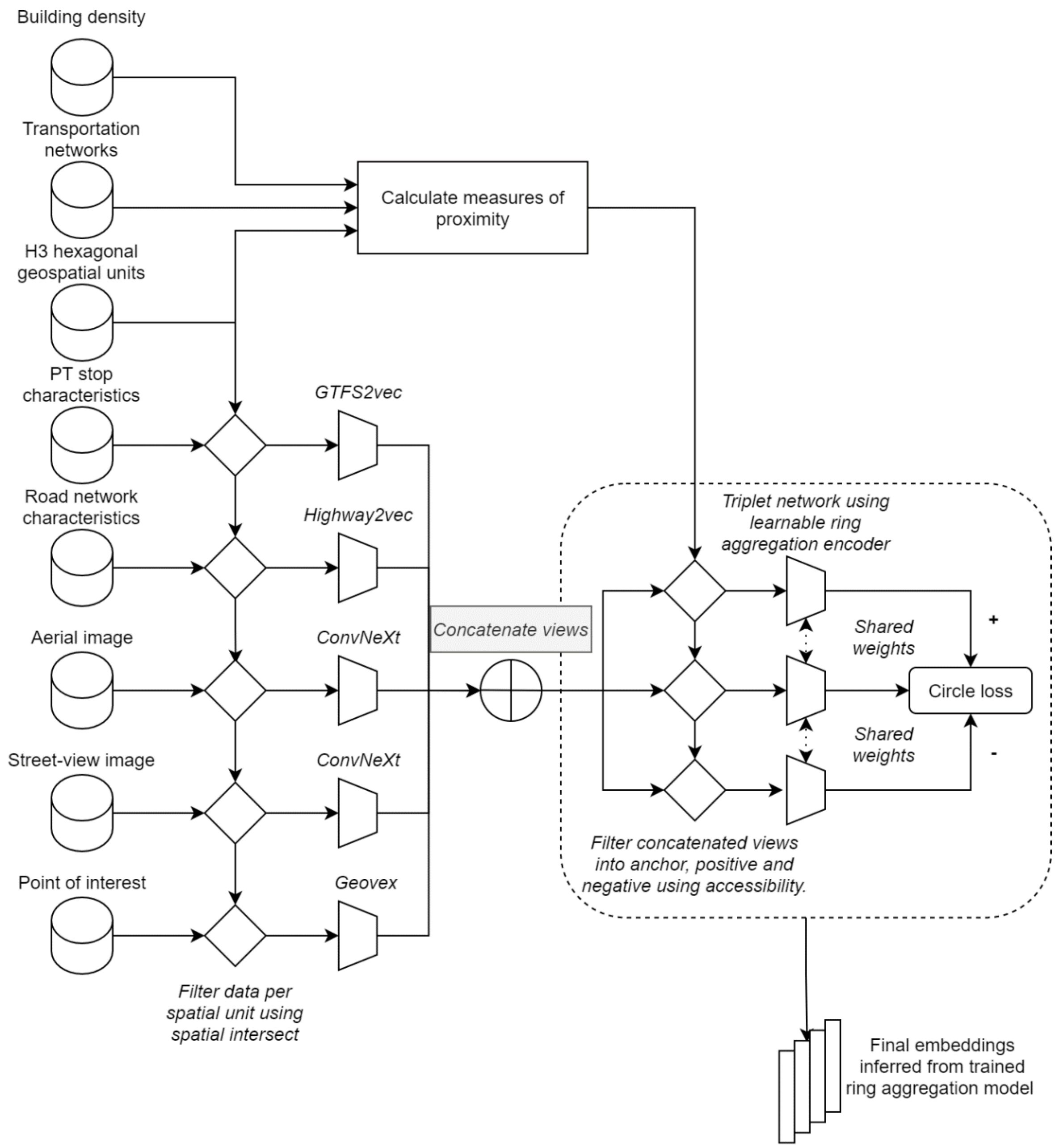


Figure 45: Complete overview of learning strategy one: a two-part process of encoding individual views and subsequent fusion.

3.2.2 Strategy Two – Sequential Similarity Loss

Learning strategy two addresses the challenge of learning across views by sequentially training encoders on the product of the previous step. The strategy consists of three distinct steps, each considering a different data modality and employing a unique sampling methodology to obtain triplets for the circle loss. The first step focuses on aerial image processing and employs a novel sampling methodology. The second step integrates point-of-interest data and uses traditional word2vec sampling. The third step refines spatial unit embeddings by considering a measure of proximity where sampling is done using weighted random walks.

1. Aerial images with 2-step ring sampling

The first step finetunes a pre-trained ConvNeXt large encoder as provided in Pytorch. Aerial images assigned to each hexagon are used to finetune the encoder and infer embeddings. Circle loss is used as similarity loss with a gamma of 250 and m of 0.25. The learning rate is set to 0.00001 with a weight decay of 0.0001 using the Adam optimiser. The sampling of triplets is done using our developed 2-step ring sampling procedure. 2-step ring sampling is a computationally efficient way to sample triplets as it does not require the retrieval of large neighbourhoods. A neighbourhood is all hexagons in the vicinity of another one. Specifically, we use the get_neighbours_at_distance function from the SRAI package to sample any k-ring and only those rings which are needed.

The sampling procedure involves two steps:

1. Sample a ring integer: positive for green (where constraints have the upper hand) or negative for red (where diffusion is strongest). For H3 resolution 9, our 2-step sampling method includes two hexagons in each ring. In contrast, at resolution 10, this increases to 4 hexagons per ring, corresponding to approximately 250 and 500 meters from the centroid of the central hexagon.
2. Unroll this ring of spatial units and sample another integer to select the spatial unit ID. The aerial images associated with the selected spatial unit IDs are then fed into the ConvNeXt encoder model for finetuning.

The 2-step ring sampling approach attempts to mimic the formulation of generalised free energy as developed by (Koudahl et al., 2023). Generalised free energy is a synthetically engineered objective function. It can be decomposed into variational free energy and negative mutual information. The former relates to achieving an objective, whereas the latter introduces an epistemic component maximising the entropy of observations. Applied to sampling spatial units, it may take the form of a tight distribution overlapped by a flat one. The former captures the achievement of an objective, whereas the latter considers the broader context. To make this dual functionality discrete, we consider the points of overlap between the distributions. Where the tighter distribution is dominant is where positive instances of the triplet should be sampled. Moreover, where the flatter distribution is dominant is where negative instances of the triplet should be sampled. Hence, the distances are 250 and 500 meters.

Compared to Hex2vec, 2-step ring sampling only considers the local context for both positive and negative samples. Contrary to hex2vec, which attempts to mimic negative sampling as done in skip-gram, which considers the entire corpus of data, in this case, all other hexagons in the study area. Hex2vec requires the wasteful retrieval of millions of spatial unit IDs.

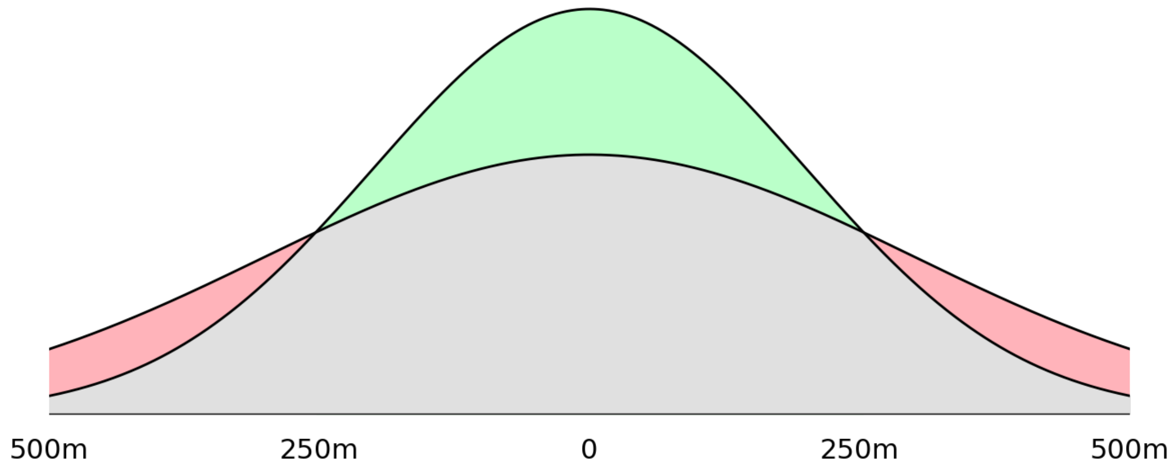


Figure 46: 2-step ring sampling derived from generalised free energy. Variational free energy, green. Negative mutual information, red.

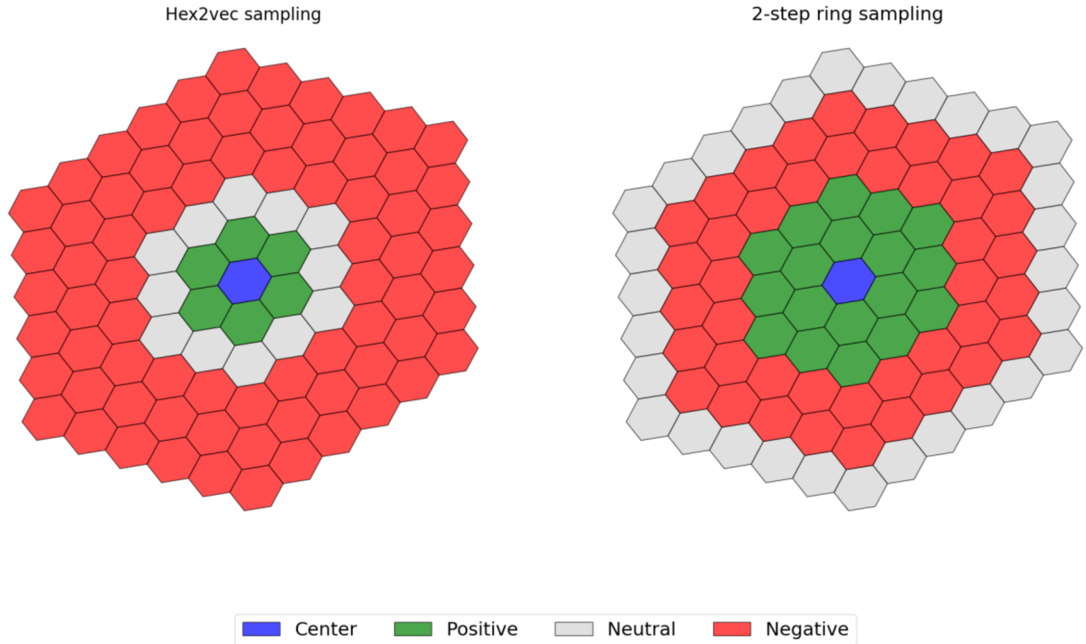


Figure 47: Comparison of the local neighbourhood in terms of anchor (centre), positive, negative and neural areas.

2. Point of interest labels with negative sampling from corpus frequencies

The second step incorporates point of interest (POI) data, employing skip-gram sampling with negative sampling from the corpus. Two shallow embedding models are used—one for POI data and another for spatial units. The dimensionality of both is 200, necessitating the application of principal component analysis to the embeddings of step one, which are dim 2048 as retrieved from ConvNeXt. Shallow embedding models are essentially dictionaries or look-up tables of weights. The spatial unit embedding model is initialised using the final weights of step one.

The anchor of triplets sampled is sampled from the spatial unit embedding model. On the other hand, positive and negative values are sampled from the POI embedding model. The dimensionality of the spatial unit embedding and POI embeddings are the same. Positive sampling queries within the same spatial unit as the anchor any POI. Negative sampling follows a power law distribution, with a power of 0.75, based on POI label frequencies calculated across all occurrences in the dataset.

A learning rate of 0.000001 and early stopping prevent catastrophic forgetting of previous steps.

3. A measure of proximity with weighted random walk

The final step has a single shallow embedding network, initialised with step two's final spatial unit embeddings. This network is trained using triplets sampled from random walks, which can be based on Euclidean distance or location-based accessibility. The random walk traverses a directed graph where edges represent increments of distance or accessibility of taking a car.

The adjacency matrix is calculated for all adjacent hexagons within five rings using a methodology similar to that used in learning strategy one. For any anchor, the origin of a walk and all spatial units traversed within that walk may be sampled as positive in the triplet. All remaining spatial units in the study area are the negative set from which to sample.

A learning rate of 0.000001 and early stopping prevent catastrophic forgetting of previous steps.

Overview

In sum, learning strategy two, as illustrated in Figure 48, is a sequential three-step process for creating urban representations. It begins with data collection and filtering into H3 hexagonal geospatial units. In Step 1, a triplet network using a ConvNeXT encoder is finetuned on aerial images, employing 2-step ring sampling for triplet sampling. Step 2 uses embeddings from the trained ConvNeXT model as anchors, alongside point of interest (POI) data, to train two shallow embedding models—one for spatial units and another for a POI label dictionary. The final third step, spatial unit embeddings, continues with the spatial unit embeddings of step two. Triplets are sampled through random walks based on location-based accessibility or Euclidean distance. All three steps use circle loss for training. The process incorporates data from aerial images, POI labels, the street network, and building density to produce the final embeddings.

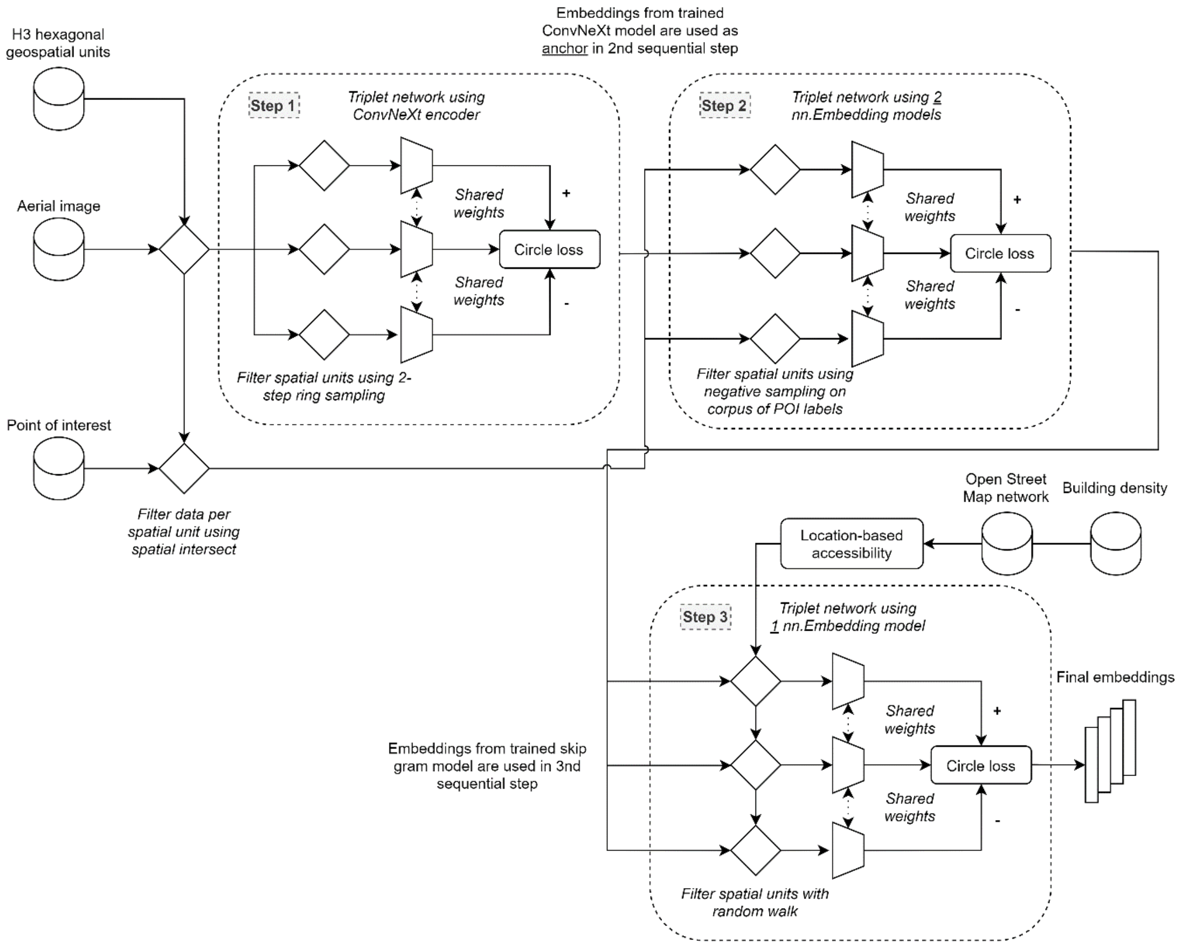


Figure 48: Complete overview of learning strategy two. Sequential three-step process. The first step is just a ConvNeXt model. The second step has a shallow embedding for both the POI dictionary and spatial units. The third only has a shallow embedding for spatial units. Note that results from previous steps are used in the circle loss of subsequent ones.

3.3 Experiments

The experiments aim to demonstrate the ability of the two learning strategies to act as transmission in a future information engine. The neural networks that create the urban embeddings are assessed using the static approach to liveability. However, instead of valuing the urban representations (percepts), they are compared to a form of apparent liveability: the Leefbaarometer score. Hence, experiments do not aim to show which configurations of urban representation methodology are most accurate as a measure of static liveability but those that show promise to be used in dynamic liveability in future work as evaluated on a static operationalisation.

Experimentation is conducted through quantitative analysis, applying the representations to two downstream tasks. First, the quality of urban representations is assessed through visualisations of agglomerative clustering. Agglomerative clustering is hierarchical, aligning well with the hierarchical structure of the H3 geospatial index. Second, the performance of urban representations is assessed through their predictive accuracy on the Leefbaarometer scores. We use the R-squared value of linear regression to assess accuracy with an underlying assumption of linearity. Assumed linearity also applies to the generous use of principal component analysis to ensure equal dimensionality (30) before applying regression.

RQ4 concerns the sampling heuristic used to train the neural network weights. Both learning strategies one and two rely on this sampling heuristic. The sampling heuristic influences which spatial units should be more or less similar, guiding the gradient descent process of the neural network used to infer representations.

RQ4 involves configuring the implemented spatial convolution. Only learning strategy one incorporates spatial convolutions. Larger receptive fields with more k-rings can account for more data in the urban region. However, there will be a trade-off given the fixed number of weights in the representations. Additionally, the type of weighted average used to aggregate across the mean representations of rings is evaluated. The Leefbaarometer similarly involves weighted averages in the construction of its indicators.

RQ5 evaluates the inclusion of different data modalities in the multi-view fusion process. It only applies to learning strategy one since learning strategy two has a fixed order of sequentially incorporating data modalities. Aggregation networks are trained using all data modalities, and testing is done by filling the remaining data sources with zeros. Learning strategies one and two are compared using aerial images and points of interest.

RQ6 involves creating and evaluating representations for all three sequential steps, including two versions for the third step, as studied in research question one.

As a bonus, all experiments cover both H3 resolutions 9 and 10. Preliminary experiments show that resolution is the highest leverage modelling decision. Experiments for resolution 10 do not cover street-view images or location-based accessibility.

Identifier	Question	Methodology
RQ3	What is the impact of the chosen proximity measure in the sampling heuristic used to calculate similarity loss?	Evaluate the triplet sampling heuristic of the fusion network: learned ring aggregation using Euclidean distance vs. location-based accessibility.
RQ4	What is the impact of configuration on aggregating over the local spatial context?	Evaluate the spatial convolution: different numbers of k-rings and weighted average types for both learned and simple ring aggregation.
RQ5	What is the added value of different data sources?	Compare the performance of models trained with various combinations of data sources.
RQ6	What is the impact of learning strategy?	Compare learning strategies one and two using aerial images and points of interest.

4 Results

This chapter covers a qualitative assessment of embeddings from individual views before addressing the research questions one by one. The underlying philosophy of the results section is that any difference not visible by rough graphs is irrelevant. After all, urban embeddings are not a precise measure of anything by themselves; they merely indicate differences between spatial units.

4.1 Individual Views

Before addressing the individual views, it is worthwhile to emphasise the drastic impact of H3 resolution. An agglomerative plot in ten clusters for the finetuned ConvNeXt model trained on aerial images using two-step ring sampling is shown in Figure 49.

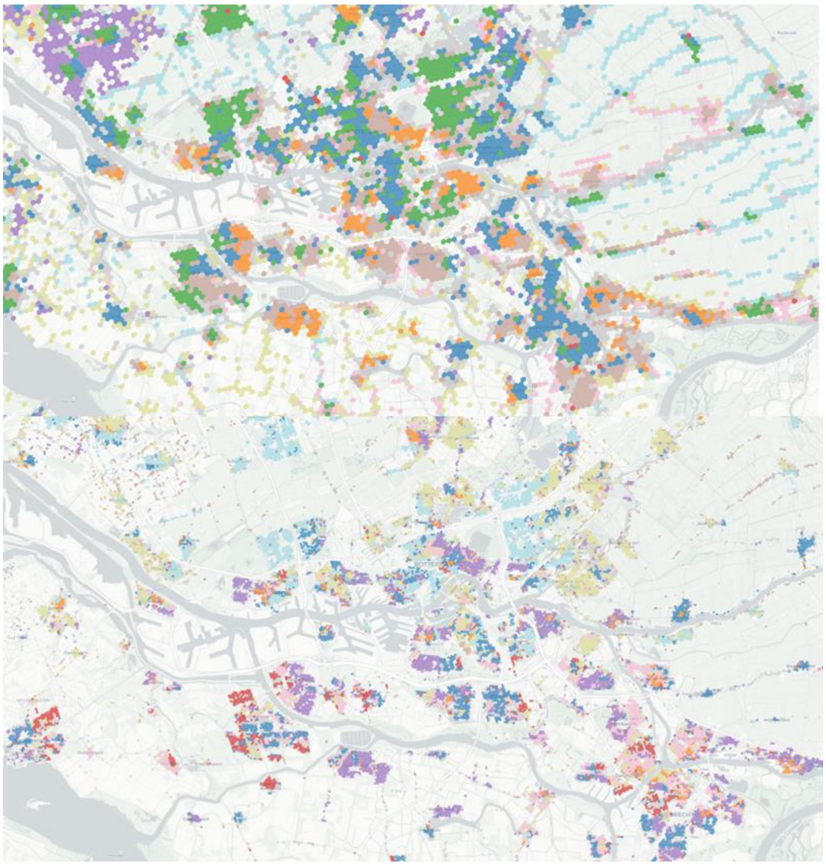


Figure 49: Comparison of embeddings from aerial images using a fine-tuned ConvNeXt model for the Rotterdam region. Top image H3 resolution 9, bottom H3 resolution 10. Colours are only distinctive of clusters within each plot.

Next is an overview of all individual views. Two point-of-interest encoders, Hex2vec and Geovex, are included to understand better how these differ in practice. Aerial images (Figure 50): Clustering shows good isolation of farmland (blue), grassland (green), greenhouses (red), dunes (yellow) and different types of urbanisation (orange and purple). Street view embeddings (Figure 53): Mirror the findings from aerial images. GTFS embeddings (Figure 51): Most pronounced in urban areas with strands connecting these. Highway2vec embeddings (Figure 54): Can differentiate highways (purple), country roads (red), urban arterials (blue) and local streets (grey). POI embeddings: Geovex (Figure 52) learns smoother representations, whereas Hex2vec (Figure 55) learns more granularly.

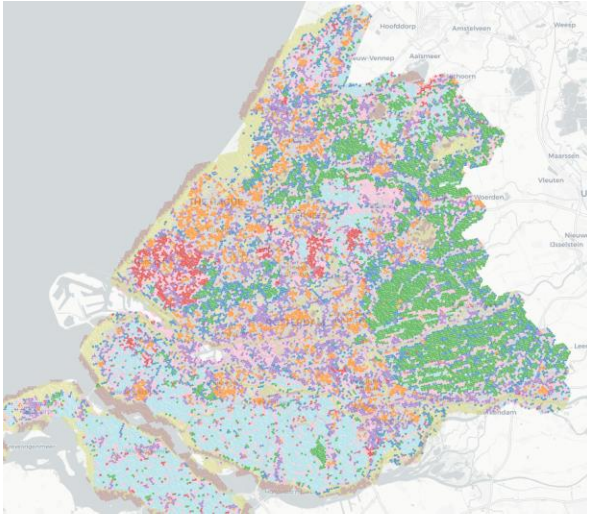


Figure 50: aerial images (pre-trained ConvNeXt model, no finetune). H3 res 9.

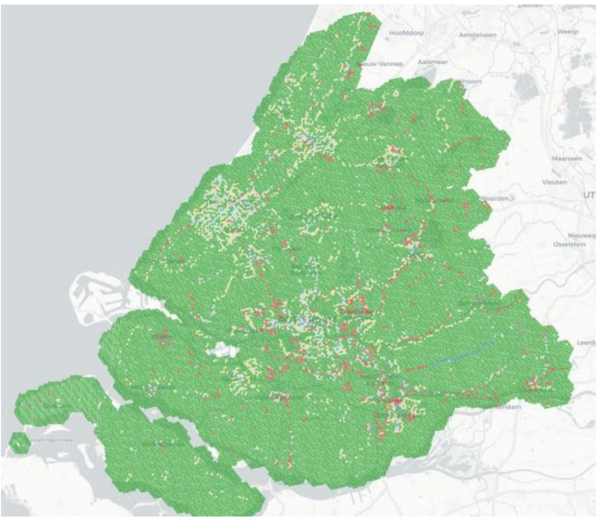


Figure 51: General transit feed specification embeddings. H3 res 9.

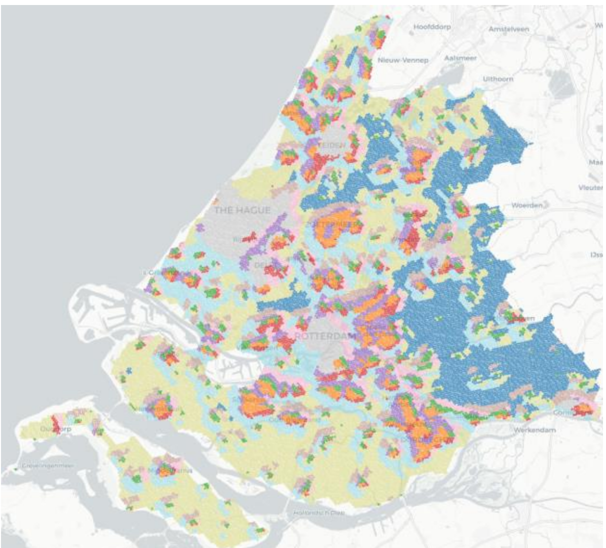


Figure 52: POI embeddings using Geovex. H3 res 9.

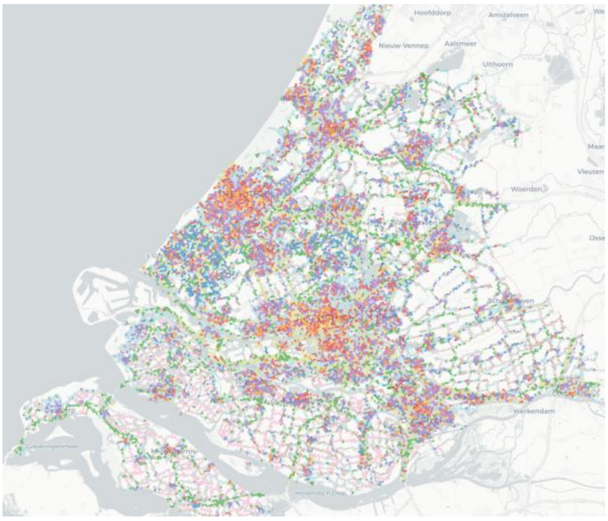


Figure 53: Street view embeddings, average per spatial unit. H3 res 9.

Figure 54: highway2vec embeddings. H3 res 9.

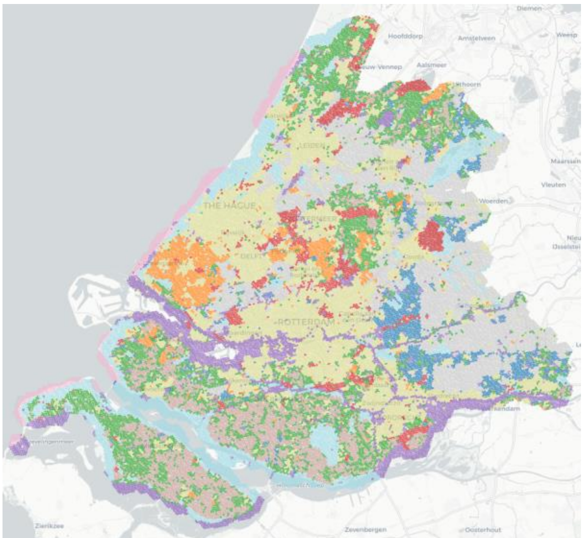


Figure 55: POI embeddings using Hex2vec. H3 res 9.

4.2 RQ3 measure of Proximity

Research question three addresses the importance of sampling heuristics in our proposed learnt aggregation network. The neural network, which aggregates over local context, is trained using triplets sampled based on the measure of proximity. For a complete comparison, we use all five data modalities (street view images, aerial images, points of interest, road network characteristics, and general transit feed specification). Furthermore, a logarithmic weighted average across five k-rings is used to configure the aggregation network. Due to memory constraints related to the size of the adjacency matrix for location-based accessibility, only H3 resolution 9 is evaluated here.

The cluster plot of embeddings created using location-based accessibility shows that the larger urban areas of The Hague (light blue) and Rotterdam (red) are kept together. When using Euclidean distance, these urban areas are broken apart (Figure 56 & Figure 57).

Quantitative results indicate that the Leefbaarometer's predictive ability differs little. However, using H3 resolution 10 instead of 9 greatly improved performance (Figure 58).

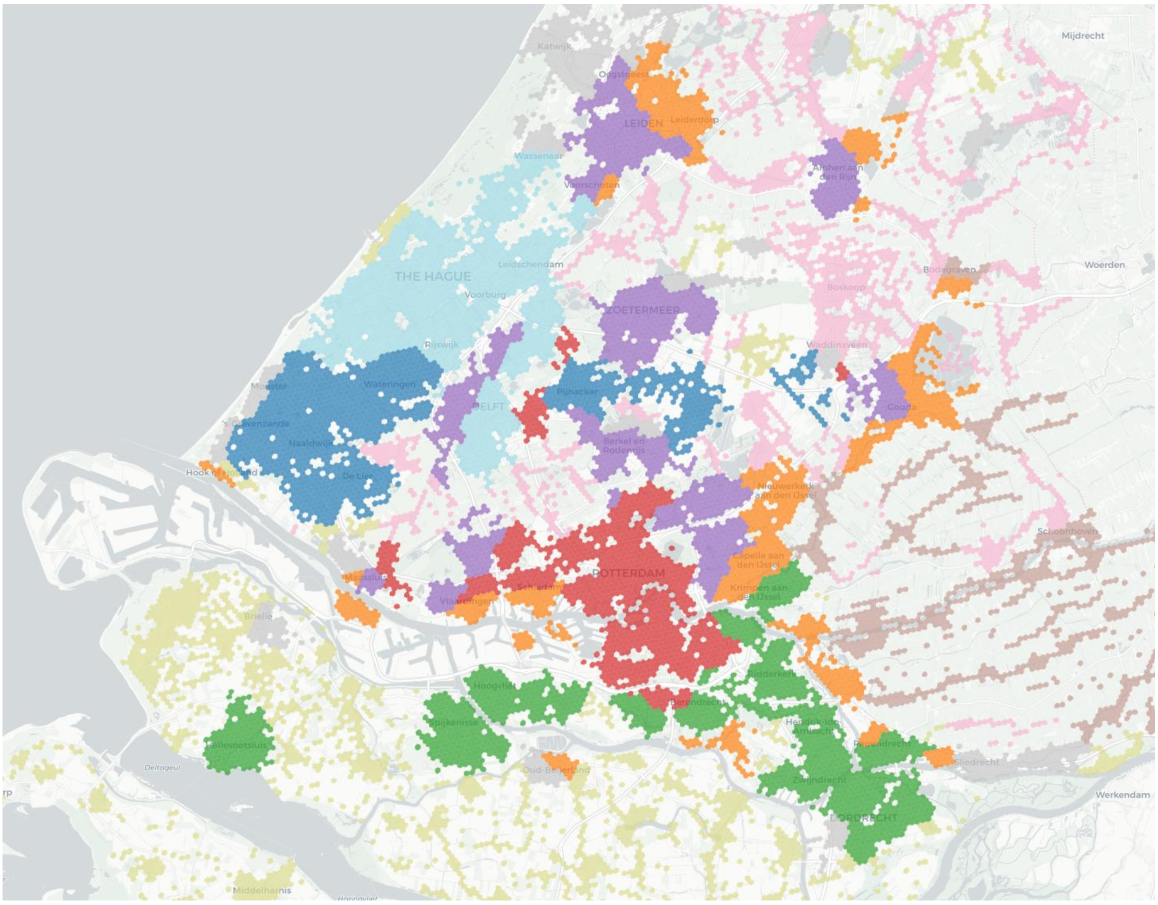


Figure 56: Agglomerative plot with all views and 10 clusters for location-based accessibility at H3 res 9.

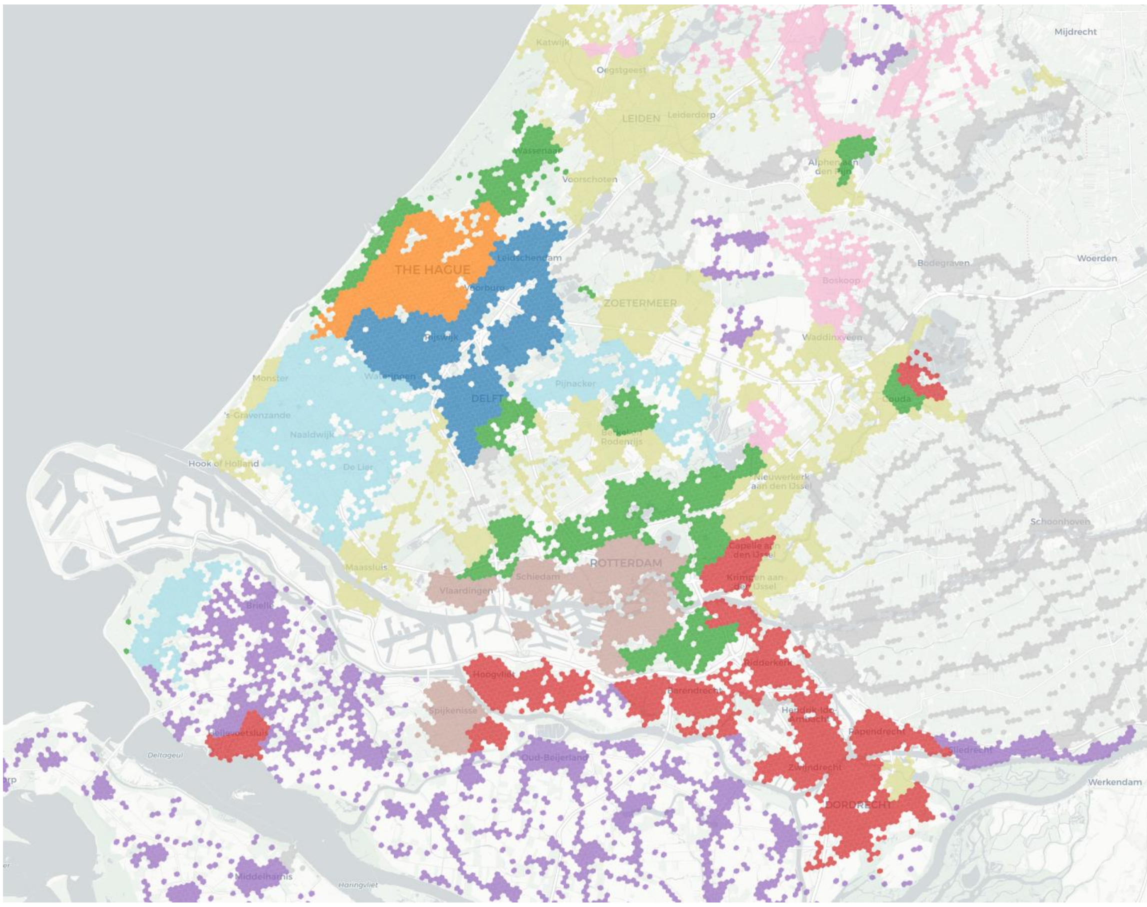


Figure 57: Agglomerative plot with all views and 10 clusters for Euclidean distance at H3 res 9.

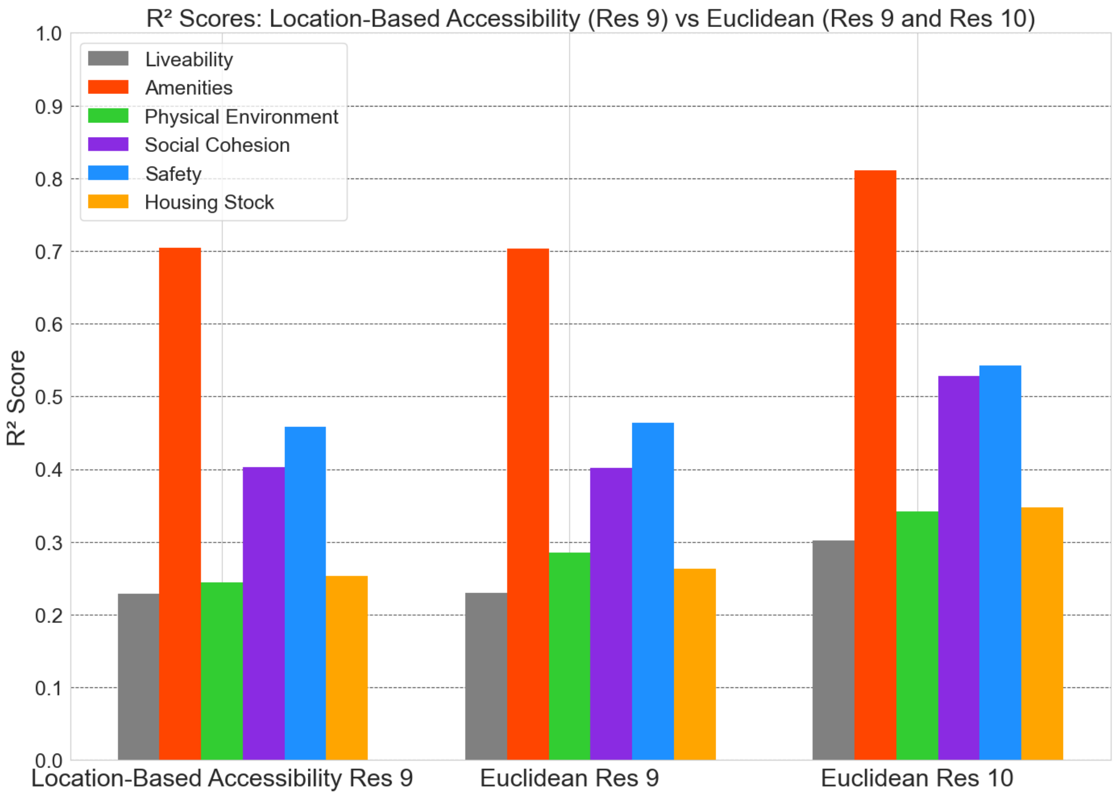


Figure 58: Comparison of explained variance between embeddings and Leefbaarometer scores. The best configurations for each are without using fine-tuned aerial images.

4.3 RQ4 Configuration of Ring Aggregation

Our proposed learning strategy applies ring aggregation in both a simple and learnt setting. The learnt model has trainable neural network weights for both within and across rings. Additionally, pooling is applied after each neural network, a mean within the ring and a weighted average across them. In simple ring aggregation, this means calculating the mean within the ring and then the weighted average without trainable neural network weights.

The impact of proximity, relevant to the learnt aggregation, has already been discussed in RQ3. Leaving two dials to configure the ring aggregation model: the number of rings to aggregate across and the weighted average used to pool across rings. While there are only two dials to configure ring aggregation, there are around twenty combinations. Hence, to get an overview of the results, we only select the best configuration, keeping the input data sources constant and using all data sources concatenated together.

Table 3 shows that a larger number of rings to aggregate across (k-rings) tends to perform better than fewer rings. A notable exception is social cohesion, which is actually best predicted using a lower k. Additionally, the best-performing k seems lower than the maximum of 15 for several Leefbaarometer scores. In terms of weighted average, most prefer exponential weighting, whereas the physical environment scores best using a logarithm.

Table 3: Best and worst configurations to predict Leefbaarometer scores. H3 resolution 10 and simple aggregation.

Target	Best k	Best Weighting	Best R^2	Worst k	Worst Weighting	Worst R^2
Liveability	5	exponential_e	0.35	1	linear	0.18
Amenities	15	flat	0.85	1	flat	0.62
Physical Environment	15	logarithm	0.40	1	flat	0.14
Social Cohesion	5	exponential_e	0.60	15	flat	0.45
Safety	10	linear	0.54	1	linear	0.39
Housing Stock	5	exponential_e	0.39	1	linear	0.25

Embeddings that score best seem to have larger clusters rather than smaller scattered clusters throughout; the number of k-rings to aggregate across likely plays a role. Additionally, there seems to be little difference between types of weighted averages for higher values of k, whereas at lower scales, this does have a great impact. For example, linear weighting assigns greater value to the centre than more distant ones and has pronounced clusters, whereas flat weighting has scattered clusters, see figures in the right column.

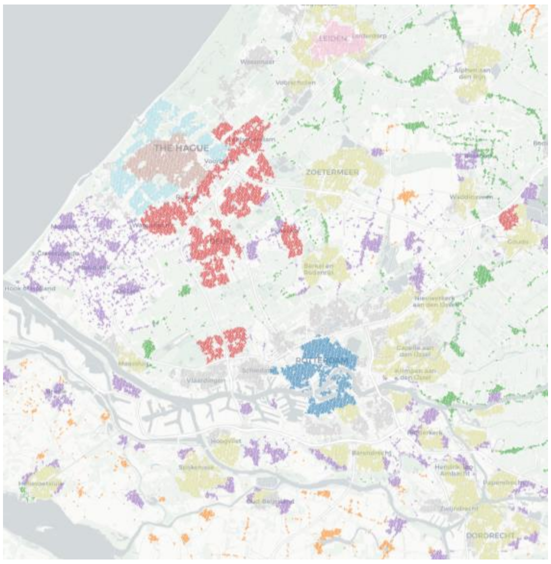


Figure 59: Liveability k=5 exponential weighting. Best.

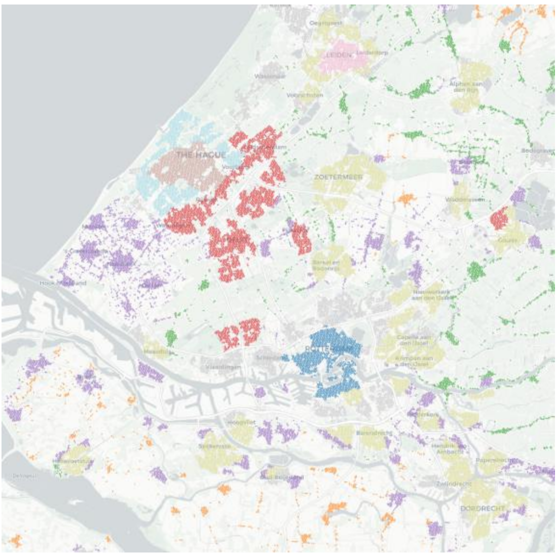


Figure 60: Amenities k=15 flat weighting. Best.

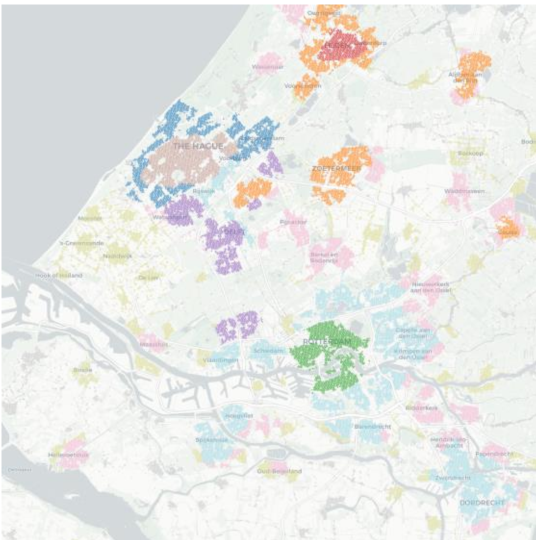


Figure 61: Physical environment k=15 logarithm. Best.

Figure 62: Liveability k=1 linear weighting. Worst.

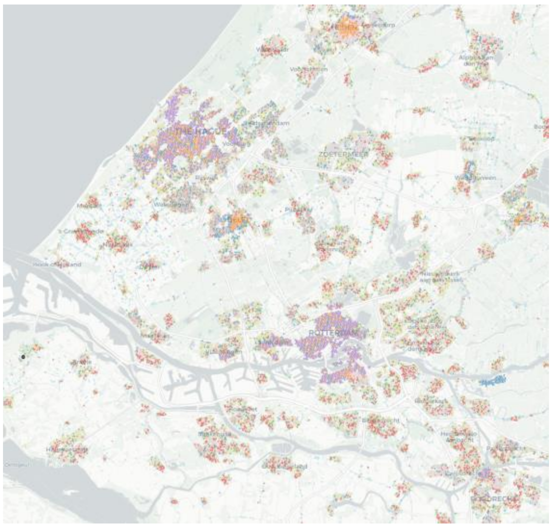


Figure 63: Amenities k=1 flat weighting. Worst.

Figure 64: Physical environment k=1 flat weighting. Worst.

4.3.1 Simple aggregation at resolution 9

Selecting the best-performing data type from all run experiments, it becomes clear that simple aggregation does not seem to be affected much by the number of rings beyond 3, after which r-squared scores start to flatten off. However, there is still significant variation between the types of weighted averages. Mostly, the 'flatter' types of averaging, like flat and logarithm, tend to improve performance with higher values of k. Additionally, when selecting the best type of weighted average, little difference can be spotted (Figure 65 & Figure 66).

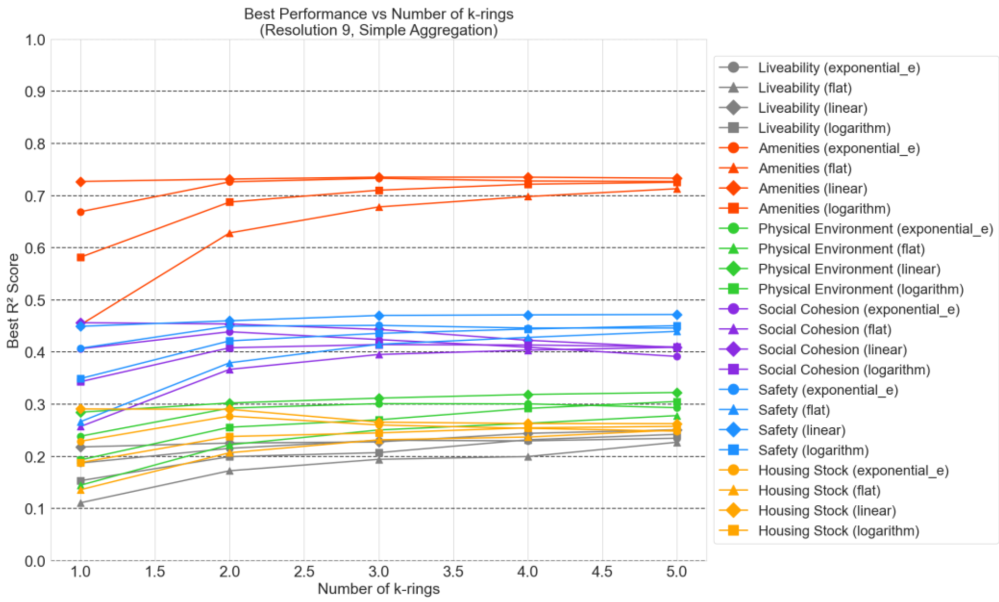


Figure 65: Predictive accuracy of Leefbaarometer score for different values of k-ring aggregated over. H3 resolution 9 simple aggregation.

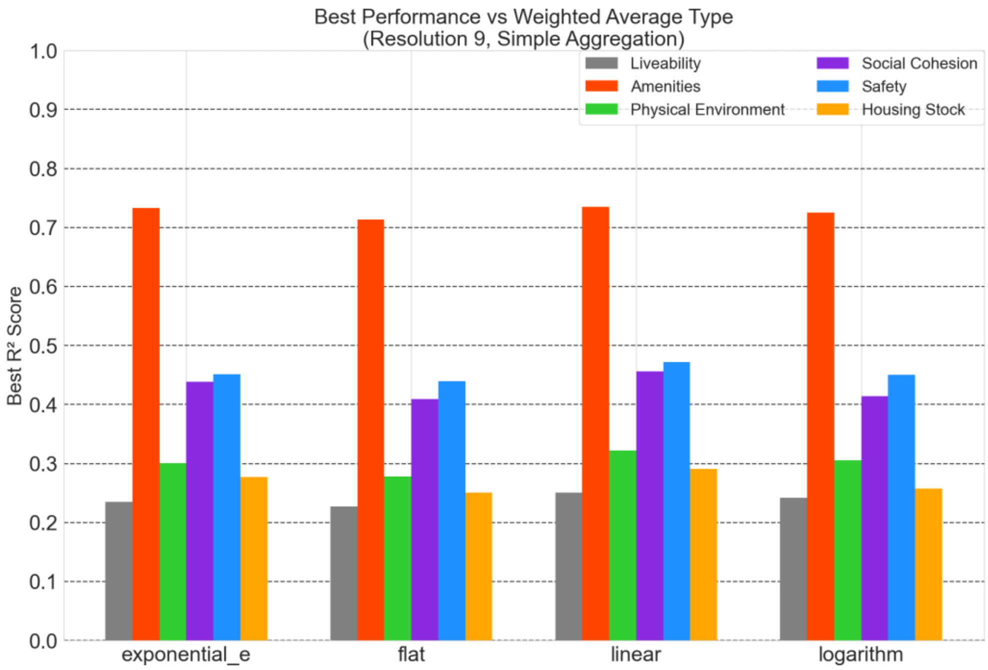


Figure 66: Predictive accuracy of Leefbaarometer scores for different types of weighted averages. H3 resolution 9 simple aggregation.

4.3.2 Learnt aggregation at resolution 9

Using Euclidean distance as a measure of proximity in the training of the learnt aggregation network appears to have little impact on the number of k-rings. Only amenities seem affected by lower values for k, whereas the performance of the other scores tends to flatten off quickly beyond about two k-rings. As for the type of weighted average, there seems to be little difference, with each type outperforming others slightly on specific Leefbaarometer scores (Figure 67 & Figure 68).

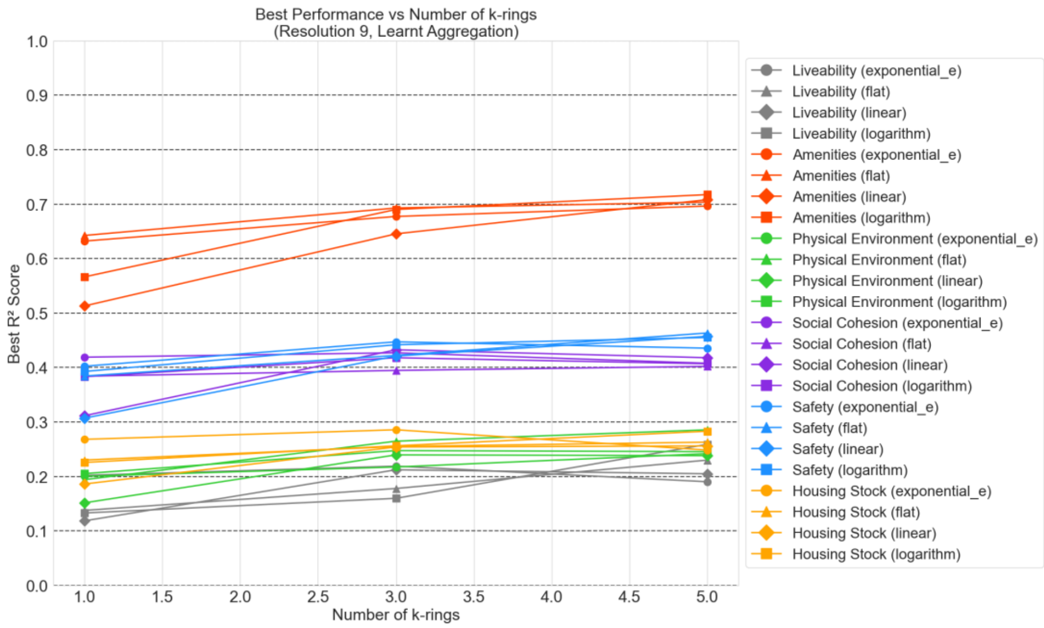


Figure 67: Predictive accuracy of Leefbaarometer score for different values of k-ring aggregated over. H3 resolution 9 learnt aggregation.

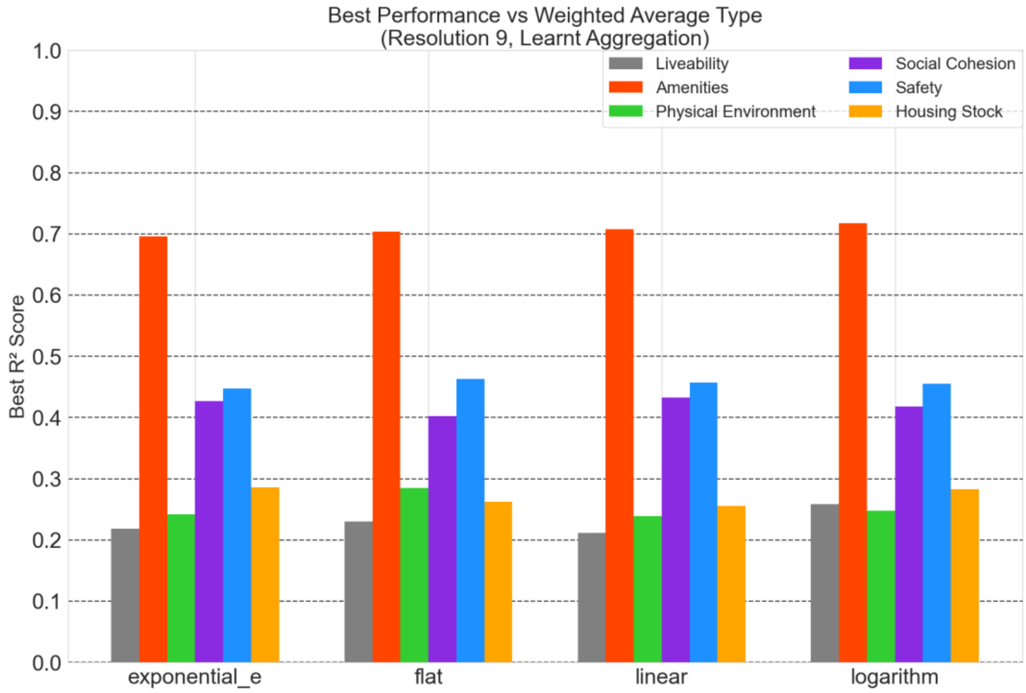


Figure 68: Predictive accuracy of Leefbaarometer scores for different types of weighted averages. H3 resolution 9 learnt aggregation.

4.3.3 Simple aggregation at resolution 10

Simple aggregation at H3 resolution 10 indicates a clear impact on the number of k-rings. Amenities seem to flatten off around ten K-rings while the physical environment keeps improving with more rings. The remainder of the scores seem to top out around a k of four, after which they diminish more or less strongly depending on the type of weighted average. Particularly, flat and logarithm weighting does not lose performance as they increase in value of k, whereas the greatest peak of performance around four is by exponential and linear weighting. The figures show a clear difference between flatter and more aggressive weighting for Leefbaarometer scores in terms of housing stock, safety, and social cohesion, in addition to the overall liveability score. On the whole, when selecting the best-performing experiment, it seems that linear weighting performs best (Figure 69 & Figure 70).

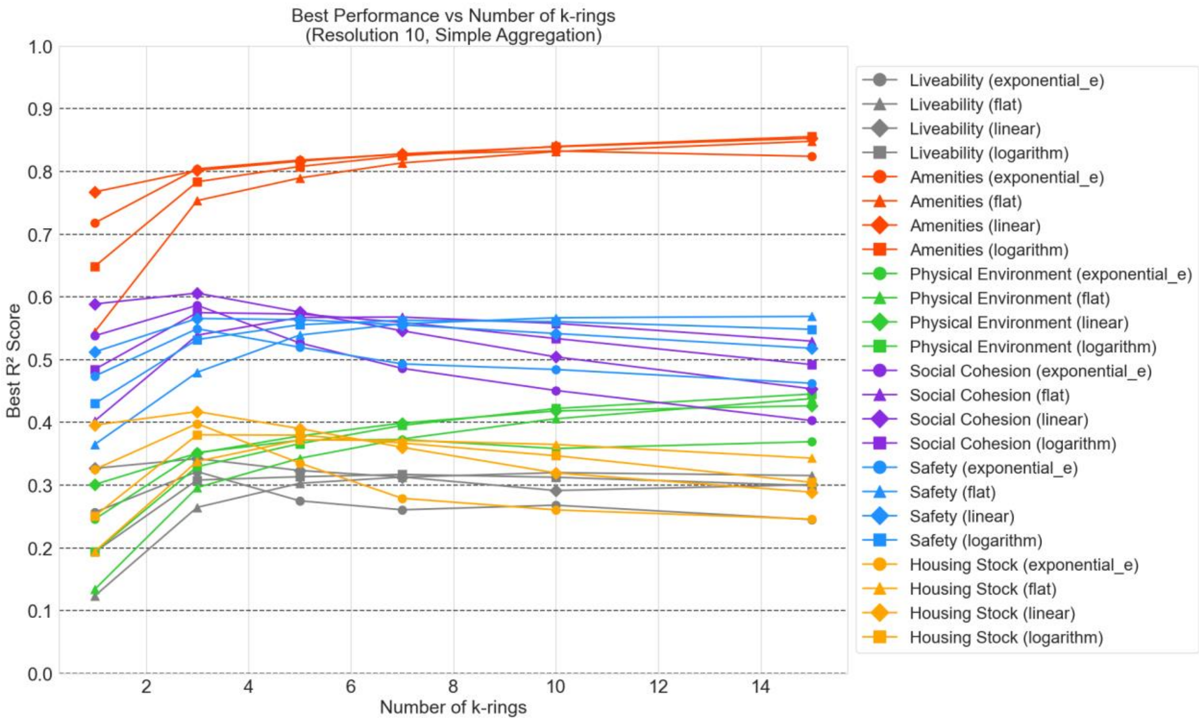


Figure 69: Predictive accuracy of Leefbaarometer score for different values of k-ring aggregated over. H3 resolution 10 simple aggregation.

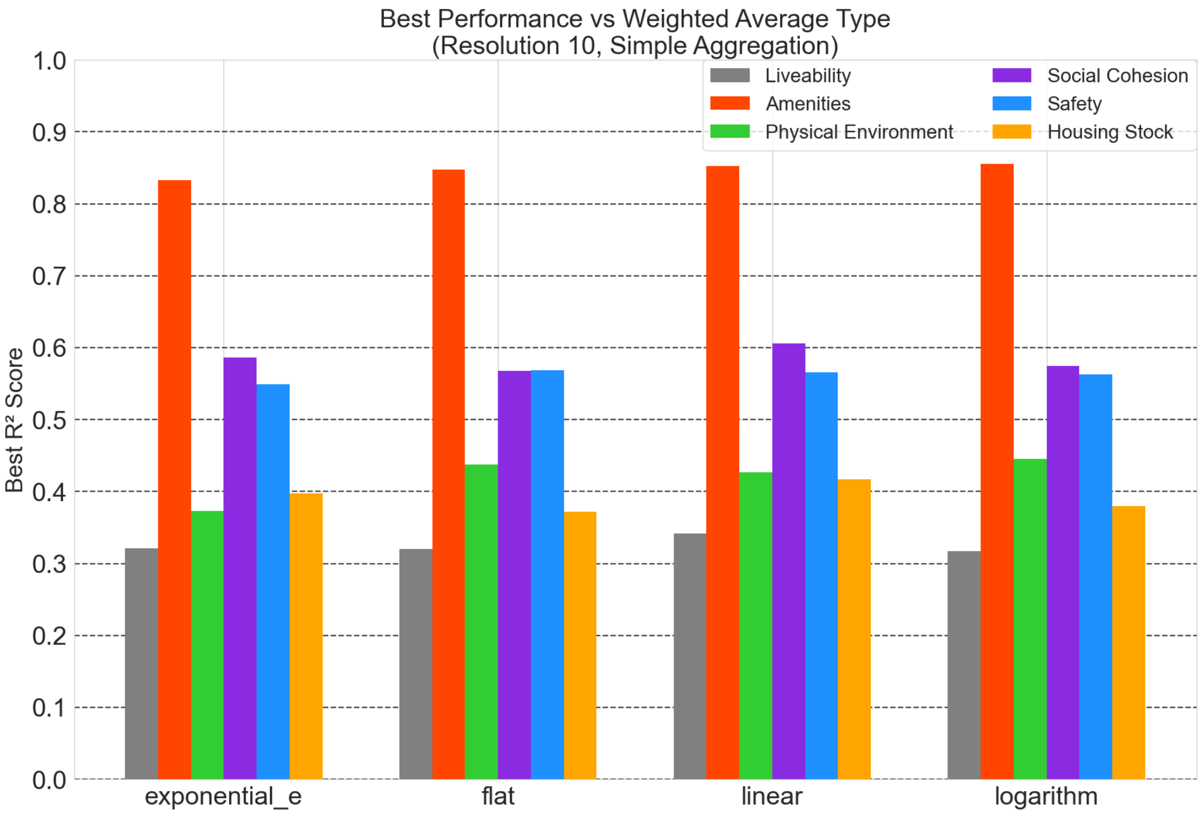


Figure 70: Predictive accuracy of Leefbaarometer scores for different types of weighted averages. H3 resolution 10 simple aggregation.

4.3.4 Learnt aggregation at resolution 10

Using Euclidean distance as a measure of proximity to train the aggregation network leads to different results at H3 resolution 10 compared to 9. Where previously predictive performance flattened off quickly, in the first quarter of the plot (note res 9 to 10 is about three-fold in k-rings for equal distance), at resolution 10, performance diminishes at larger values of k for several types of weighted average. Whereas there was a clear pattern of flatter (flat, logarithm) versus steeper types of average (exponential, linear) at resolution 9, this is not the case here. Instead, it seems to depend greatly on the specific Leefbaarometer score at hand. For example, performance on social cohesion stays relatively constant using exponential weighting, whereas the other weightings do not come close. Safety, on the other hand, performs better with linear weighting. On the whole, exponential weighting performs best for social cohesion and housing stock, whereas amenities and physical environment seem not to do so. The other three weighted averages perform roughly equally (Figure 71 & Figure 72).

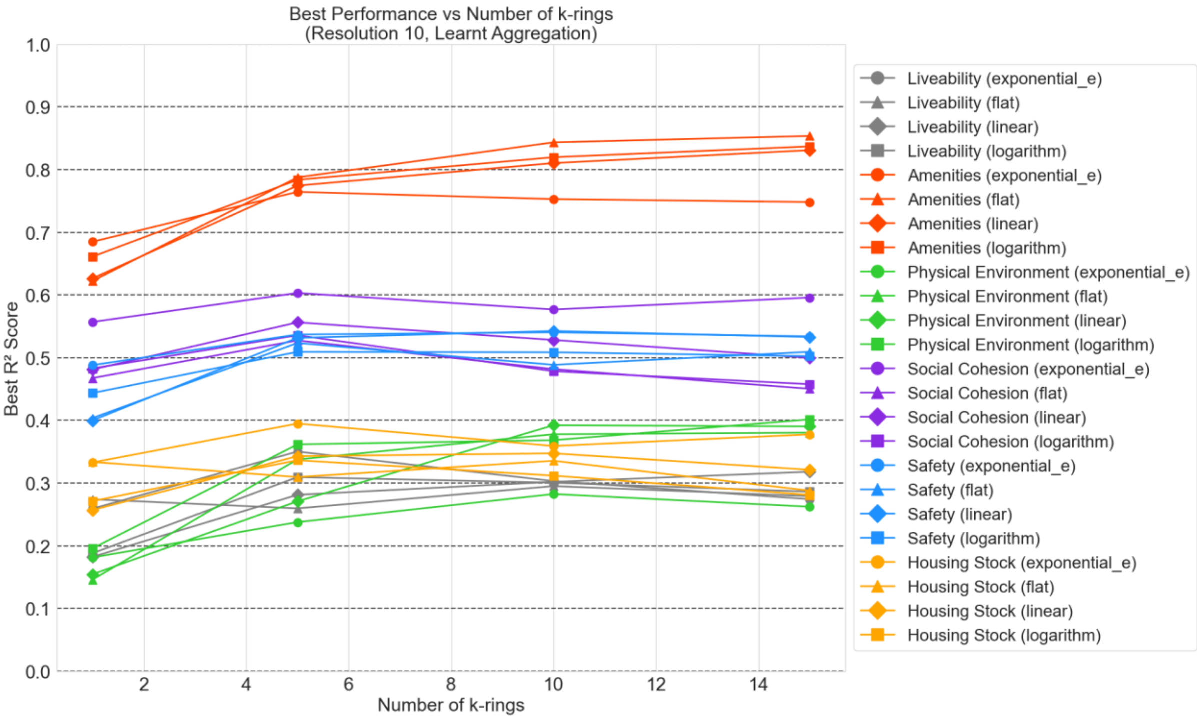


Figure 71: Predictive accuracy of Leefbaarometer score for different values of k-ring aggregated over. H3 resolution 10 learnt aggregation.

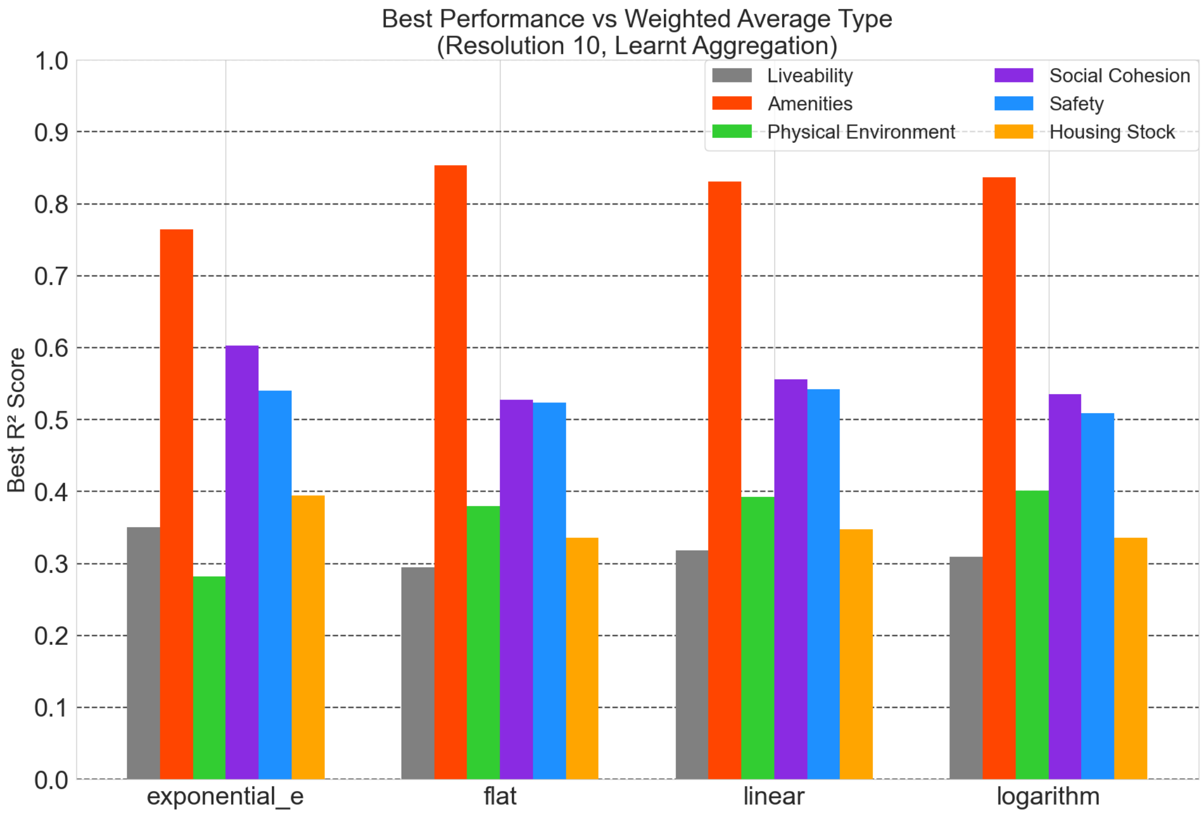


Figure 72: Predictive accuracy of Leefbaarometer scores for different types of weighted averages. H3 resolution 10 learnt aggregation.

4.4 RQ5 Data Sources

Looking at the results for simple aggregation, images and points of interest are highly informative, showing high predictive performance across the board. Within the images, there is great variation: embeddings from aerial images obtained using a fine-tuned encoder by far outperform all other modalities on liveability, social cohesion, housing stock and safety while underperforming on physical environment and amenities. Embeddings from street view and aerial images obtained without finetuning have a well-balanced performance on Leefbaarometer scores compared to others. The point of interest encoders both do well, trading off liveability and the physical environment. Finally, road network and GTFS embeddings do well in amenities, safety, and social cohesion and somewhat on housing stock, but they fall short on physical environment and liveability. Learnt aggregation performs worse across the board, particularly affecting aerial image embeddings (Figure 73 & Figure 74).

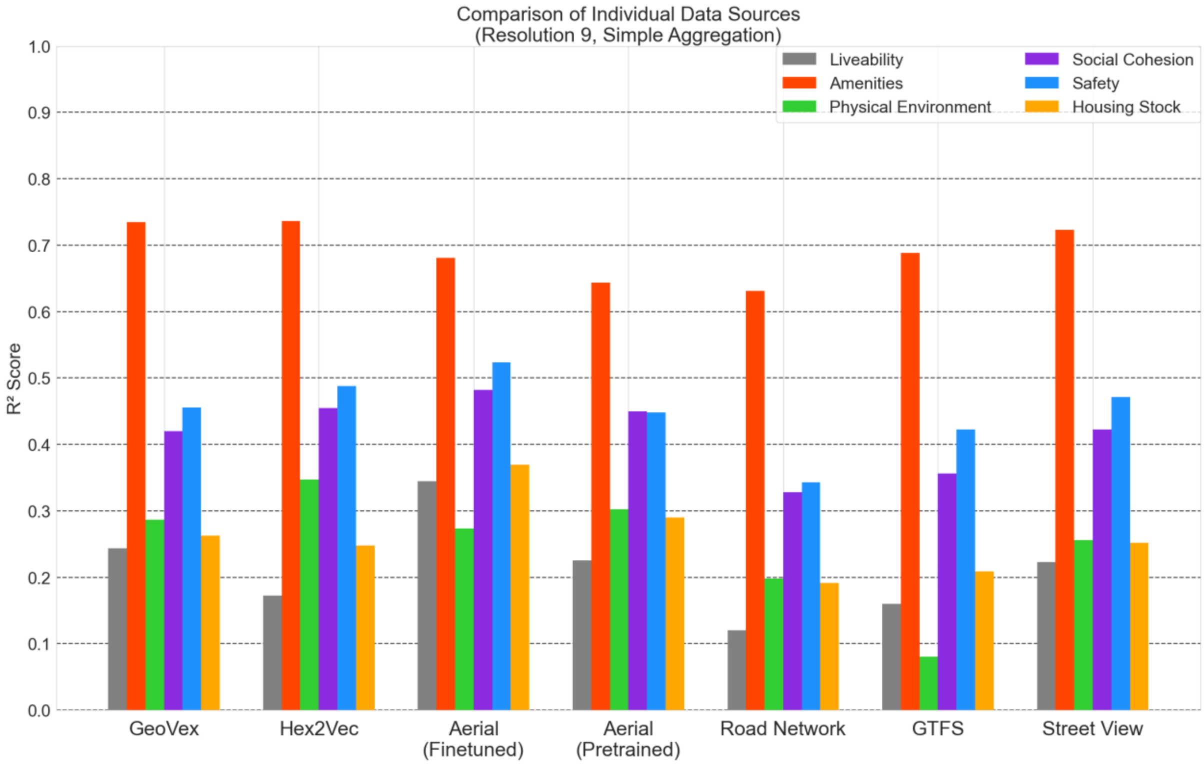


Figure 73: Performance on Leefbaarometer scores for different data sources. H3 resolution 9 simple aggregation.

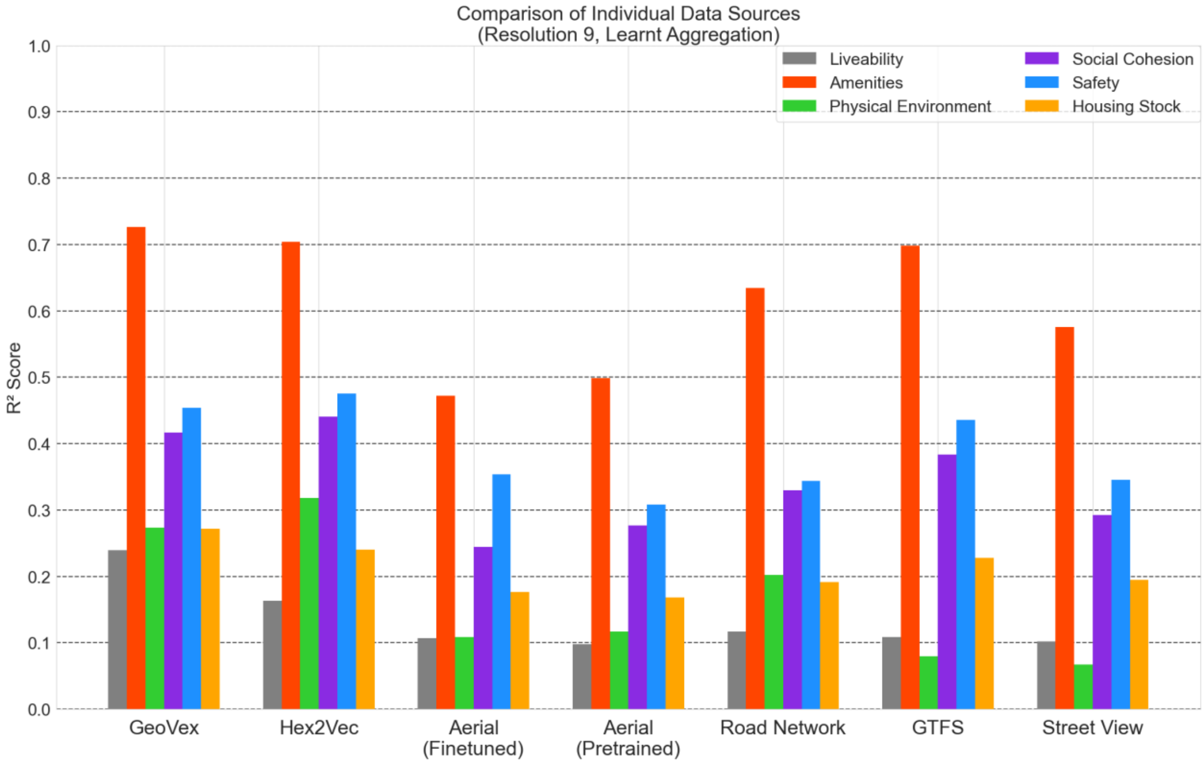


Figure 74: Performance on Leefbaarometer for different data sources using learnt aggregation (Euclidean distance as a measure of proximity). H3 resolution 9 learnt aggregation.

Considering the synergy of data sources when using all of them, it is clear that hex2vec seems to outperform Geovex at H3 resolution 9, and finetuning image encoding models does not matter much. However, there seems to be no real difference for the learnt aggregation for pre-trained or finetuned models, nor for Hex2vec or Geovex (Figure 75 & Figure 76).

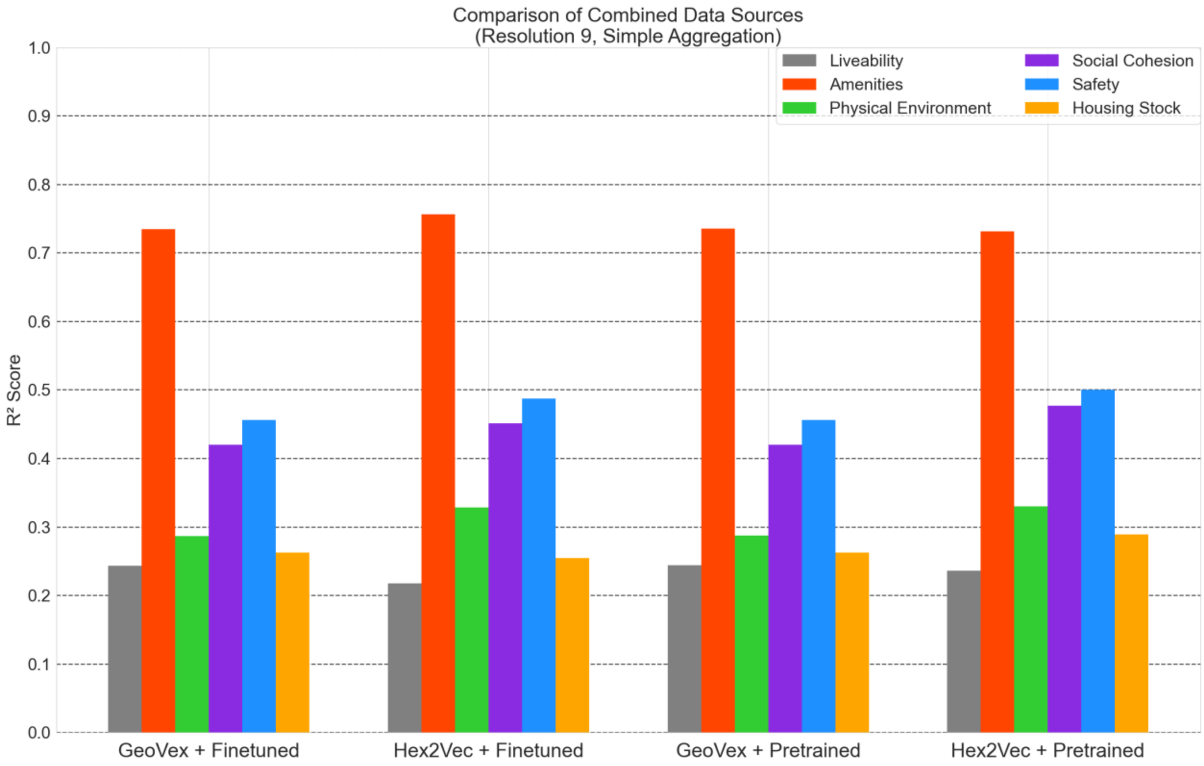


Figure 75: Comparison of predictive performance on Leefbaarometer scores using a combination of all views. H3 resolution 9 simple aggregation.

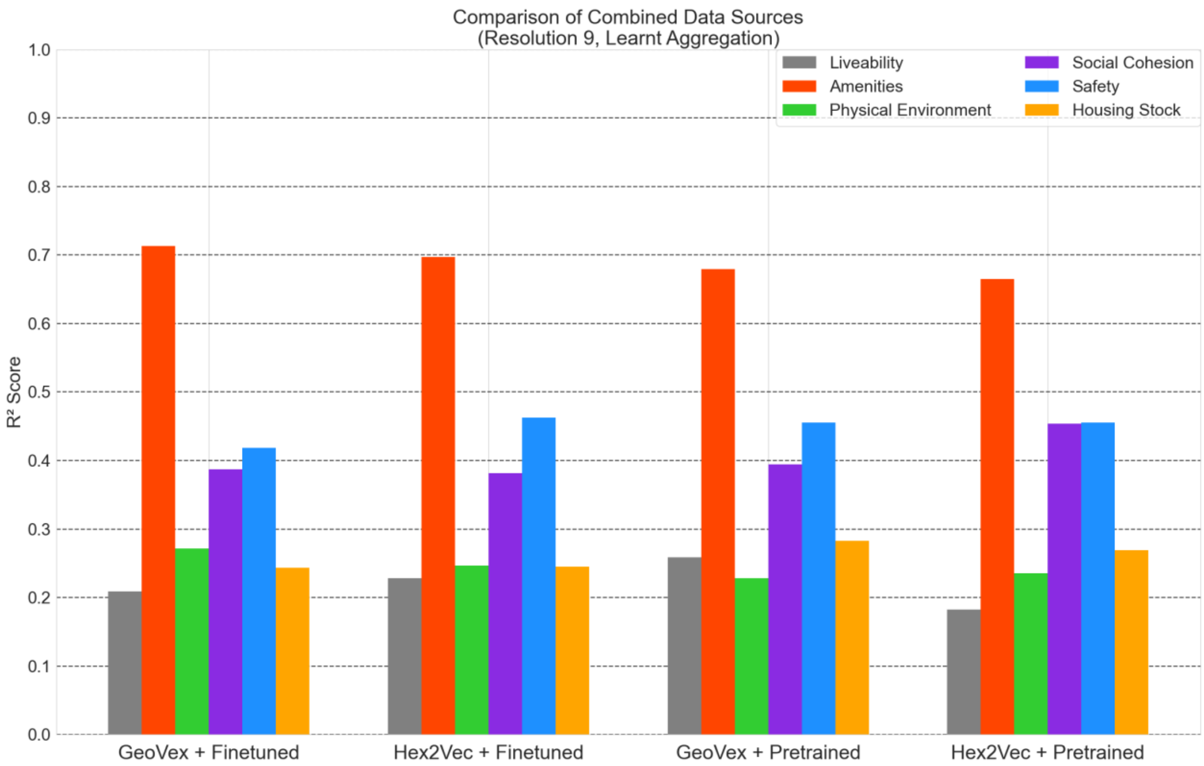


Figure 76: Comparison of predictive performance on Leefbaarometer scores using a combination of all views. H3 resolution 9 learnt aggregation.

4.5 RQ6 Learning Strategy Two

Learning strategy two involves three steps, each building upon the embeddings of spatial units from the previous ones. In turn, the predictive performance is expected to improve with every step. The greatest improvements are made in amenities, social cohesion and safety. Euclidean distance and accessibility as measures of proximity seem to perform roughly equally. The increase in resolution does seem to improve predictive performance in terms of amenities, social cohesion and safety; however, this is not the case for the physical environment (Figure 77 & Figure 78). Inspection of cluster plots shows that embeddings become more refined from steps one to two, forming cohesive islands. Step three shows drastic changes by focusing purely on proximity. Comparing Euclidean to accessibility shows that the seaside is clustered with the Hague and Zoetermeer with the 'Groene Hart' when accounting for accessibility.

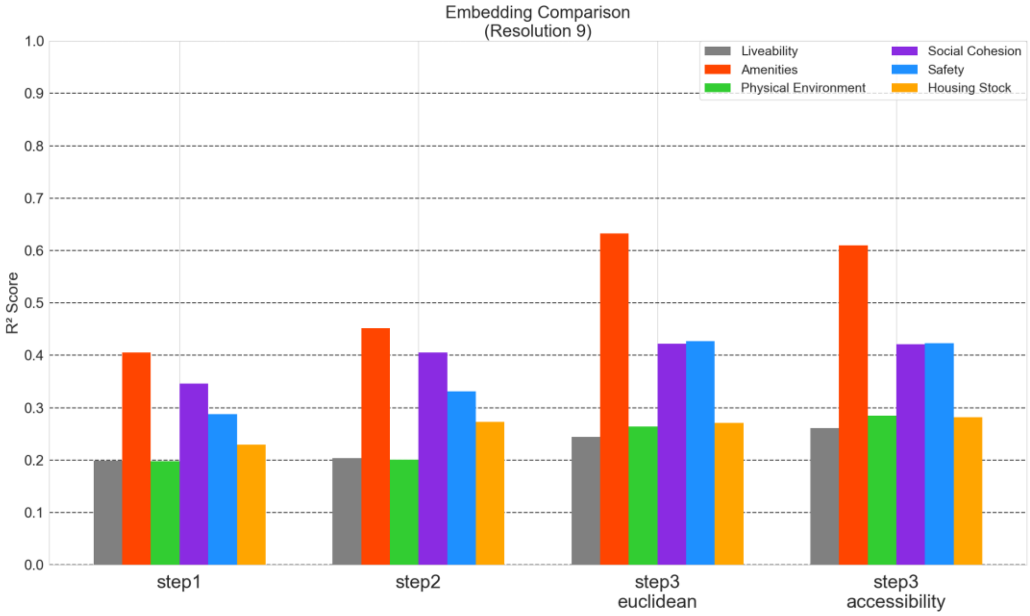


Figure 77: Learning strategy two - sequential integration of data sources comparison at H3 resolution 9.

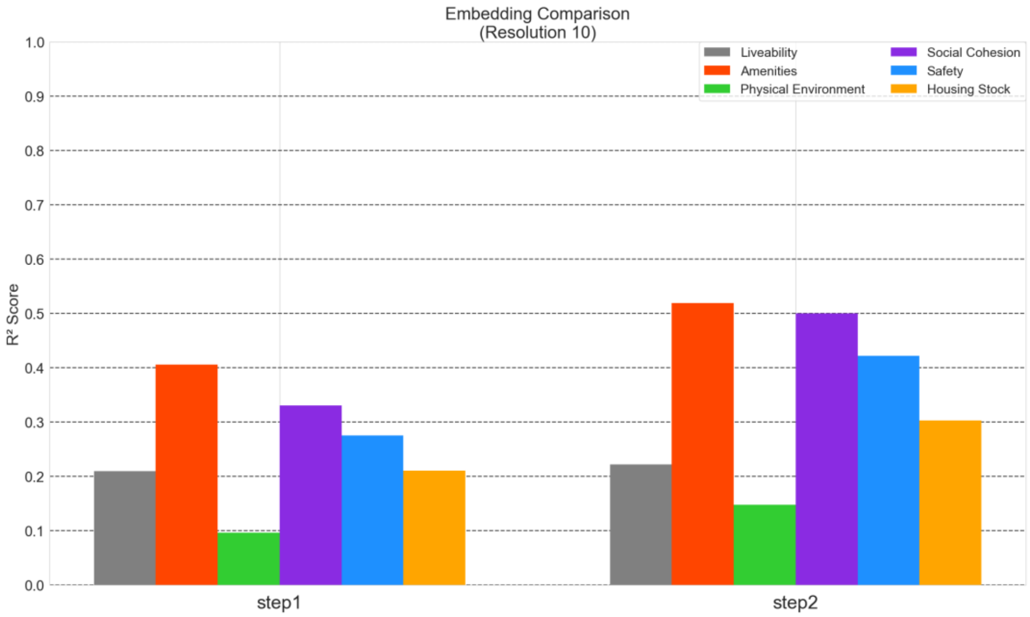


Figure 78: Learning strategy two - sequential integration of data sources comparison at H3 resolution 10.

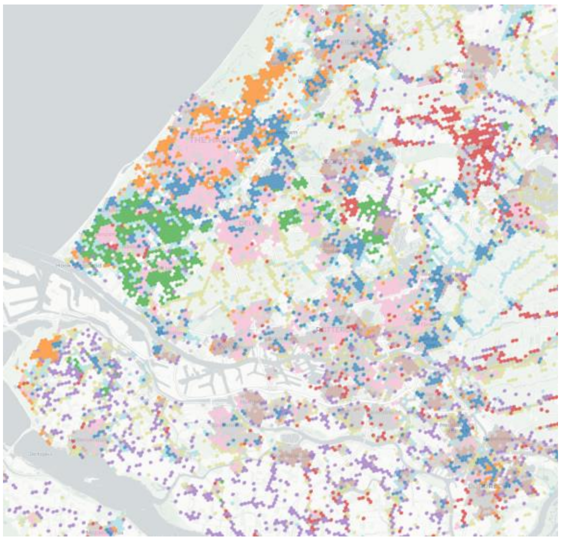


Figure 79: Resolution 9 - step 1 - aerial images.

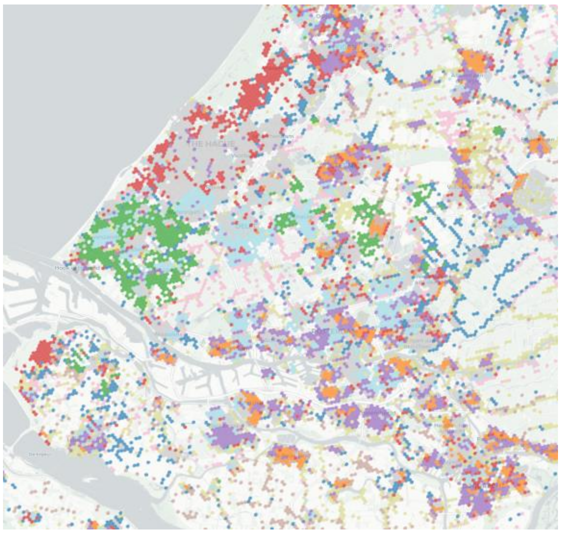


Figure 80: Resolution 9 - step 2 - Point of Interest.

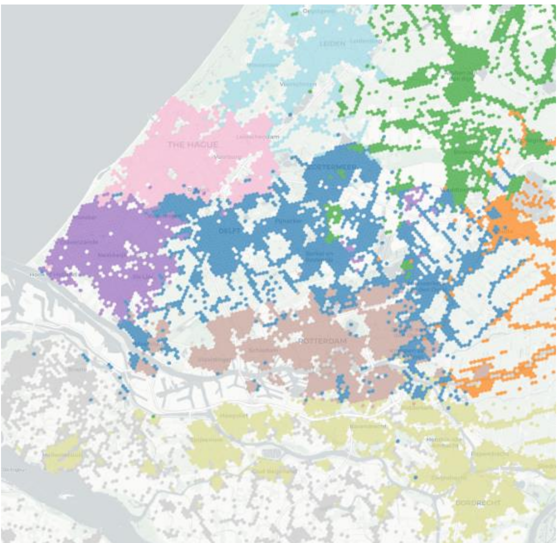


Figure 81: Resolution 9 - step 3 - Euclidean.

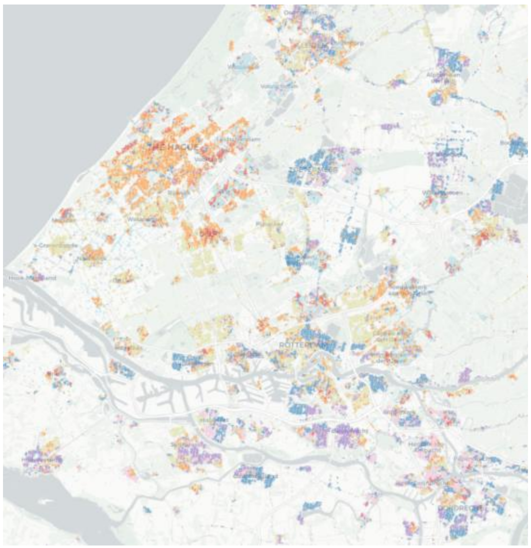


Figure 82: Resolution 10 - step 1 - aerial images.

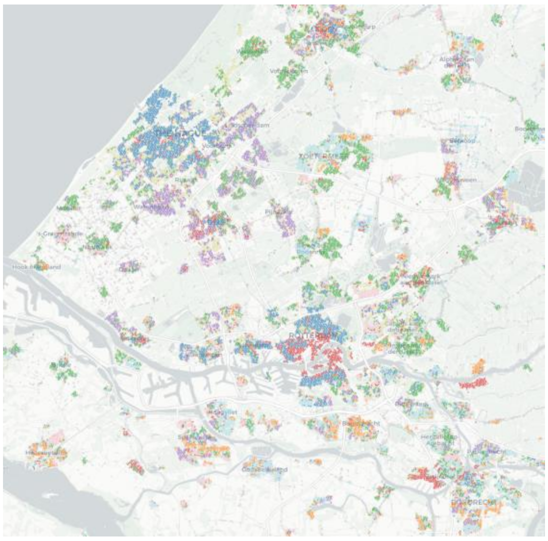


Figure 83: Resolution 10 - step 2 - Point of Interest.

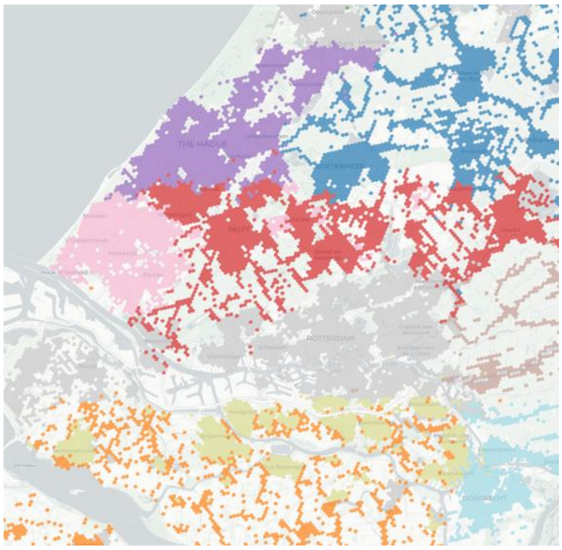


Figure 84: Resolution 9 - step 3 - Location Based Accessibility

5 Conclusion

This thesis set out to understand the role of representations in defining and operationalising liveability. The research objective is tackled by developing a theoretical framework and a modelling study that builds upon it. A 2x2 matrix captures the resultant theoretical framework by delineating the operationalisation and definition of liveability for the static and dynamic approaches. Static liveability is the current state of the art (Leefbaarometer), whereas dynamic liveability draws upon the active inference literature.

The modelling study subsequently attempts to bridge these two approaches, bootstrapping the development of dynamic operationalisations. The dynamic approach involves information engines which construct and maintain niches. Urban representation learning is applied to craft and experiment on the configuration of a transmission for such an engine. In this study, empirical urban representations are compared to Leefbaarometer scores. Of which there are five: Safety, Social Cohesion, Housing Stock, Amenities, and Physical Environment.

Safety is measured using objective indicators like crime rates and subjective indicators such as perceived safety. Social Cohesion considers factors like demographic diversity and community engagement. Housing Stock evaluates aspects such as vacancy rates and building characteristics. Amenities assess the proximity to various services and job accessibility. Physical Environment includes indicators like green space, environmental risks, and urban heat.

The research questions tackle the research objective in two steps, first addressing the definition and then the operationalisation. Since the definition influences the operationalisation, the report is similarly structured. First, the definition is understood by covering the relationship between liveability and representations split into two questions: The role of the action-perception loop in liveability and the role representations play in the action-perception loop. Second, operationalisation builds upon the definition outlined in the theoretical framework and concerns the development of a transmission using urban representation learning. Four research questions address the development and configuration of the transmission. A new learning strategy is developed, taking full advantage of hexagonal spatial units (H3). Learning strategy one fuses different data modalities and approximates spatial convolutions.

Identifier	Question	Methodology
RQ1	What is the role of the action-perception loop in liveability?	Literature Review
RQ2	What is the role of representations in the action-perception loop?	Literature Review
RQ3	What is the impact of the chosen proximity measure in the sampling heuristic used to calculate similarity loss?	Modelling Study
RQ4	What is the impact of configuration on aggregating over the local spatial context?	Modelling Study
RQ5	What is the added value of different data sources?	Modelling Study
RQ6	What is the impact of learning strategy?	Modelling Study

5.1 What is the role of the action-perception loop in liveability?

The action-perception loop plays a distinct role in static and dynamic approaches to liveability, as illustrated in the conceptual model. The conceptual model in Figure 85 highlights the key difference in how these approaches conceptualise the fit between residents and their environment. Overall, the action-perception loop only applies to the dynamic approach of liveability. The static approach to liveability is a linear measurement of outcomes rather than an iterative looping process. The two approaches ultimately reflect how we view decision-making, driven by rational choice or the construction of narratives.

In the static approach, liveability is viewed as an outcome of a linear process. Indicators are mapped to percepts, which then interact with needs and desires. Perception in this view is aligned with passive observation, a bottom-up construction process of stimuli into percepts. The fit between environment and resident occurs between perception and needs/desires. It suggests that liveability is a static outcome resulting from the valuation of how well subjective percepts of the environment meet the residents' needs/desires.

The dynamic approach presents an interactive model. Indicators interact with percepts biased towards the characteristic state of the niche embodied by the resident, where characteristic states are described by the needs/desires. The location of fit implies that it is an iterative, continuous process of palpating those indicators which align with the biased percepts. In turn, percepts interact amongst themselves as they mirror the dynamics of the environment, and they are only occasionally informed of these dynamics by measurement of indicators. The bidirectional arrow contains the top-down generative (←) and bottom-up recognition densities (→) that define active inference. Niche construction is the process of configuring the generative density to reduce the divergence between it and the recognition density. Needs/desires bias the niche construction process towards characteristic states—characterising the niche.

The relevance of niche construction to transport policy is found in residential self-selection and idea flow. Residential self-selection is the phenomenon in which travel behaviour is explained away through either geographic characteristics or personal travel preferences. Controlling for personal travel preferences leads to a zero loading on geographic characteristics in the structural equation model. Hence, residents who prefer certain travel affordances will have moved their house to a niche satisfying these. On the other hand, idea flow relates to social niche construction as studied in extended active inference. Belief sharing and synchronisation is the object of interest and process to model.

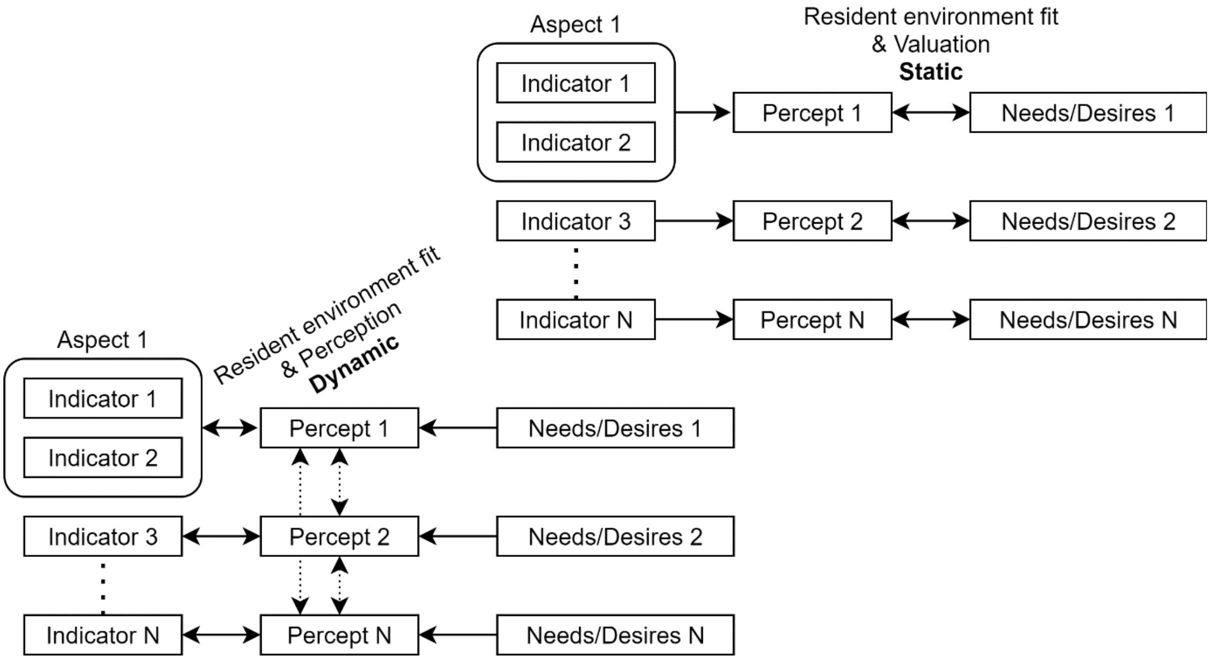


Figure 85: Conceptual differences between the static and dynamic approach to liveability.

5.2 What is the role of representations in the action-perception loop?

Representations are instrumental in the action-perception loop. From an active inference perspective, they parameterise the internal states of the generative model and are continuously updated to reflect interaction with external states given the blanket.

As the name suggests, the generative model generates impressions on the blanket from within. It palpates indicators for their expected values. The system will move when these indicators are active states, such as actuators or muscles. In turn, movement is itself an observation to be made—self-evidencing. Representations are only instrumental in informing the selection of optimal actions, leading to new observations. Therefore, perception is an action, as observations are purposefully palpated based on pragmatic and epistemic values.

The demarcation between static and dynamic approaches to liveability becomes even more crisp when considering the role of representations. Static liveability considers representations to be fixed outcomes and more or less objectively representative of the environment, as required for rational choice. Dynamic liveability takes a pragmatic turn and aligns itself with an enactive view in which representations are merely instrumental in selecting optimal actions. Representations enable action selection beyond mere habituated reflexes, the latter sometimes confused to be definitive of enactivism. The combination of habitual reflex and contemplative reason through representations defines enactivism and the dynamic approach.

5.3 What is the impact of the chosen proximity measure in the sampling heuristic used to calculate similarity loss?

Two measures of proximity are evaluated: location-based accessibility and Euclidean distance. Empirical results indicate little difference in predictive accuracy on Leefbaarometer scores between these two. However, agglomerative clustering revealed that location-based accessibility tends to keep larger urban areas intact, such as The Hague and Rotterdam.

It additionally captures highway-adjacent suburbs. Euclidean distance breaks them apart and appears to focus on the physical environment, as evidenced by the higher accuracy of this Leefbaarometer score. Higher H3 resolutions improve predictive performance more than the measure of proximity, indicating the importance of spatial granularity in urban representations.

5.4 What is the impact of configuration on aggregating over the local spatial context?

The configuration of ring aggregation significantly impacts the quality of urban representations. A larger context or receptive field generally performs better, with some exceptions, such as social cohesion. Exponential weighting is preferred for most Leefbaarometer scores, while physical environment scores are best when using logarithmic weighting. At H3 resolution 9, simple aggregation is minimally affected by rings beyond 3, while learned aggregation shows little impact from the number of k-rings except for amenities. At H3 resolution 10, there is a clear impact of k-rings, with different Leefbaarometer scores peaking at different k values. The type of weighted average (flat, logarithmic, exponential, linear) has varying impacts depending on the specific Leefbaarometer score and resolution. Flatter weighting (flat, logarithm) peaks later, while steeper weighting (exponential, linear) earlier. Indicators targeting the resident side of the fit between resident and environment, see Figure 7, tend to perform better with steeper weightings (social cohesion, safety, liveability, housing stock), whereas indicators approximating the physical environment do better with flatter weightings (physical environment, amenities).

5.5 What is the added value of different data sources?

Images (aerial and street view) and points of interest are highly informative, showing high predictive performance across Leefbaarometer scores. Finetuned aerial images outperform other modalities for liveability, social cohesion, housing stock, and safety but underperform for the physical environment and amenities. Street view and non-finetuned aerial images show well-balanced performance across Leefbaarometer scores. Point-of-interest encoders perform well, trading between liveability and physical environment scores (Geovex, Hex2vec). Road network and GTFS embeddings excel in predicting amenities, safety, and social cohesion but lag in the physical environment and liveability predictions. When using all data sources, those with Hex2vec outperform combinations with Geovex at H3 resolution 9, while finetuning image models has little impact on performance.

5.6 What is the impact of learning strategy?

Learning strategy two involves a sequential three-step process based on similarity loss, showing stepwise improvements in social cohesion and safety accuracy. In step three, where the sampling heuristic is based on proximity, Euclidean distance and accessibility perform similarly. Higher resolution (H3 resolution 10) improves performance for amenities, social cohesion, and safety, but not for the physical environment. Furthermore, our proposed learning strategy one with ring aggregation outperforms learning strategy two, particularly at higher resolutions. Cluster plots show embeddings become more refined from steps one to two, forming cohesive islands, while step three shows drastic changes by focusing purely on proximity.

6 Discussion

This thesis set out to understand the role of representations in defining and operationalising liveability. The research objective was tackled by developing a theoretical framework and a modelling study. A 2x2 matrix captures the resultant theoretical framework by delineating the operationalisation and definition of liveability for the static and dynamic approaches. Static liveability is the current state of the art (Leefbaarometer), whereas dynamic liveability draws upon the active inference literature.

The modelling study attempted to bridge these two approaches, bootstrapping the development of dynamic operationalisations. The dynamic approach involves information engines which produce work by constructing and maintaining niches. Urban representation learning was applied to craft and experiment on the configuration of a transmission for such an engine. In this study, empirical urban representations are compared to Leefbaarometer scores.

The discussion focuses on delineating these two approaches in further detail, highlighting how their complementarity is an incredible advantage when operationalising the dynamic approach to liveability. The obstacle to operationalising the dynamic approach lies in developing the correct transmission and identifying the associated characteristic states. Respectively, it involves identifying the correct set of indicators, mapping them to percepts and meeting the suitable needs/desires of the corresponding niche.

Working towards these future research directions requires an in-depth understanding of the intersection between transport modelling and active inference—drawing extensively upon the chosen H3 geospatial index and potential sampling procedures. The hexagonal spatial index used in this thesis aligns with renormalising generative models poised to become a mainstream approach to generative AI.

6.1 Rethinking Liveability: Static and Dynamic Approaches

The concept of liveability in transport policy extends far beyond traditional metrics of travel time and congestion. It encompasses the broader impact of transport infrastructure on urban life, including its influence on social interactions, economic opportunities, and environmental quality. For instance, the design of public transport networks not only affects commute times but also shapes urban form, influences land use patterns, and impacts social equity. The static approach to liveability, as exemplified by tools like the Leefbaarometer, has been instrumental in quantifying these impacts. It allows transport planners to assess how infrastructure projects might affect various outcome indicators (Huibregtse, 2021). However, this approach has limitations in capturing the dynamic nature of urban systems. Consider, for example, the introduction of a new light rail system. While static models might predict trend-wise changes in accessibility, they may not fully account for long-term shifts in residential patterns, business locations, or evolving travel behaviours that emerge as residents adapt to and interact with the new infrastructure. As such, the dynamic approach to liveability may align with the sustainable mobility paradigm, which focuses on scenario-based planning (Banister, 2008).

At first, this thesis started with the intuition that liveability is more than collecting ever larger volumes of indicators. Such a practice tends to balloon as adding one more indicator becomes relatively insignificant. Maintaining dozens of indicators becomes expensive as many labour hours must be dedicated to keeping up to date with theory, data collection and validation.

More importantly, the way transport policy evaluation looks at the world reflects those represented by the evaluation methodology. Residents do not live in a world of indicators.

Nor do travellers live in a world of rational choice, one fully mapped into distinct objects, each bestowed with a fixed set of attributes to evaluate. Policy evaluation is the study of scaling behaviour. Therefore, looking for a description of behaviour that also accounts for mental processes makes sense (Dietrich & List, 2016). Narrative is yet another explanation for the driving force behind behaviour (Bouizegarene et al., 2024), as has been described in transport policy, referring to it as *homo narrans* (Schwanen, 2020).

The ecological approach to liveability aligns with the recommendation by Leidelmeijer to take the ecological understanding of liveability seriously when developing its operationalisation. The dynamic approach to liveability is an attempt to combine disenchantment with current practice hyperfocused on indicators while fully embracing an ecological interpretation of liveability. The key differentiator between the static and dynamic approaches is the notion of niche construction, which precludes passive rational choice with a fully observed world. The niche describes the coupled reciprocal generative model between external and internal states conditional on blanket states. Internal states parameterised by representations can only be informed by that which is perceived through palpation of the external world.

Fortunately, it was possible to consolidate the dynamic nature while retaining indicators in the developed conceptual models. Indicators in the dynamic view are not just measurement outcomes presented on a silver platter but anticipated outcomes for which the resident has made preparations. For any observation on the blanket to inform internal states using the recognition density, there has to be a prior expectation, a generative density. Indicators, as used by modellers, point towards something out in the world that is of interest. It is subsequently a matter of subjectivity how the individual perceives this indicator, biased by their characteristic states better known as needs/desires.

Since indicators point towards something in the world, they may also be obtained by processing high-dimensional data using neural networks. Thus, not only is it possible to use neural networks to map indicators to percepts, but indicators themselves can also be created by neural networks. Representational indicators are justified when mapping data to representations is unambiguous, one-to-one rather than many-to-many. The underlying rationale is that low ambiguity does not require active data sampling; optimising for both pragmatic and information value is useless if there is no ambiguity to resolve using that information. Parr et al. (2024) demonstrated this by adding a cost term to sampling; at some point, the gain in information is insufficient to warrant the sampling action, and the process stops.

6.2 The Role of Representations in Liveability

Underlying the difference between the static and dynamic approaches to liveability is a fundamental concern for the role of representations. This concern is not limited to liveability but extends to artificial intelligence in general. Pezzulo et al. (2024) discuss the development of generative artificial intelligence, contrasting what may be understood as the static and dynamic approaches. Pezullo et al. conclude that the dynamic approach (active inference) experiences the results of its actions as feedback, in contrast to the static approach (e.g. ChatGPT), which only learns from what is in the training set. The dynamic approach actively samples actions that provide epistemic affordances to get feedback on uncertain parts of the world. The benefit of active sampling based on affordances is a reduction in data and computational requirements and a grounded understanding of the world rather than one inferred from second-hand sources of experience.

As the name suggests, the (active inference) generative model generates impressions on the blanket from within. Impressions are generated by sampling from the joint probability distribution, which defines the generative model. Applied to liveability, it palpates indicators for their expected values. The system will move when these indicators are active states, such as actuators or muscles. In turn, movement is itself an observation to be made. It is to see the world move and the environment pass by as one walks through it. Observing the results of one's action is called self-evidencing, such that the concept of a self is isolated from all that is not self through palpation.

Representations are instrumental in the action-perception loop. From an active inference perspective, they parameterise the internal states of the generative model and are continuously updated to reflect interaction with external states given the blanket (M. J. D. Ramstead et al., 2024). Specifically, internal states are statistical moments of probability distributions like the mean and variance. Representations are only instrumental in informing the selection of optimal actions, which generate impressions on the blanket, the divergence between impression and sensory stimuli leading to a recognition of what is. Therefore, perception is an action, as observations are purposefully palpated based on pragmatic and epistemic values and selected using an approximate generative model fitted to (historical) data by maximising accuracy and minimising complexity—looking to the past and the future simultaneously or as Albarracin, et al. (2023) frame it, retention and protention.

6.3 Urban Representation Learning: Insights and Challenges

Urban representation learning aims to capture urban environments' complex, multifaceted nature in a compact, machine-readable format. Our study leveraged various data sources and a novel learning strategy to create urban representations that could predict Leefbaarometer scores. This approach provided insights into the static measure of liveability and laid the groundwork for more dynamic, process-oriented models of urban systems.

The urban representation learning study yielded several key insights:

- Weighted averages play a crucial role in predicting Leefbaarometer scores, with exponential functions best-describing liveability, social cohesion, and safety.
- The optimal number of rings for aggregation varies across different aspects of liveability.
- Point of interest embeddings showed unexpected strength in predicting Leefbaarometer scores across all categories.
- Learnt aggregation underperformed compared to simple aggregation, likely due to the nature of the Leefbaarometer's methodology and the bottleneck effect.
- The public transport network should be included in the calculation of accessibility. Moreover, more attention should be paid to preparing the network graph.

Predicting Leefbaarometer Scores

Overall, the R-squared scores for several Leefbaarometer scores were rather low. This can be explained by considering the indicators used to craft these scores. For the physical environment, we did not account for disaster risk, wind turbines, heat stress or pollution. We did not account for isolation quality, foundation quality, ownership or rental, overcrowding or the demographics of housing stock. Finally, the overall liveability, which combines the five Leefbaarometer scores using valuation, was also remarkably unexplained by urban embeddings. The variance introduced by valuation is not found in the embeddings. However, the 2-step ring sampling methodology, in combination with aerial images using H3 resolution 10, performed all right. Nevertheless, it took much effort, so it may be safe to say that engineering the embedding process of environmental characteristics will not give high predictive accuracy on apparent liveability as operationalised using the Leerbaarometer.

The role of weighted averages in predicting Leefbaarometer scores provides new insights. At first, it was unclear whether an exponential function would best describe liveability, social cohesion, and safety. The physical environment, however, lined up with expectations, performing best using a logarithm, as described in the report detailing the development of the Leefbaarometer (Mandemakers et al., 2021). The optimal number of rings (distance) to aggregate over is not always the maximum one can afford. Maximising the number of rings is only worthwhile for amenities and the physical environment. The other scores performed better with smaller receptive fields. Results also indicate that the weighted average and optimal receptive field are coupled. Flatter functions do better with larger receptive fields, whereas steeper functions (like exponential) do better with smaller ones. The Leefbaarometer scores, which perform better with steeper functions, also tend to involve subjective indicators or liveability valuation. Of all the six scores, the more subjective ones align with steeper functions. As such, it may be worthwhile to study the impact of weighted averages in future work. At least for the static approach to liveability.

The impact of different data sources was lower than expected. All data sources were reduced in dimensionality to equal size using principal component analysis. Despite this, image data was expected to have variance to spare such that even after reduction, its embeddings would perform better. However, this did not seem to be the case. Instead, it was the point of interest embeddings which exceeded expectations—performing relatively well across all Leefbaarometer scores.

Amenities, in particular, seem insensitive to the type of data used to create embeddings. For other data sources, there are significant gains in using one source over another. Even fine-tuned aerial images do not perform best on all fronts while far exceeding expectations in predicting housing stock and liveability. Surprisingly, when using all data sources, hex2vec slightly outperforms Geovex on all scores except liveability.

Learnt aggregation performs much worse than simple aggregation, likely due to the nature of the Leefbaarometer, which aligns with simple aggregation methodology. The Leefbaarometer does not consider the spatial context to the same extent as our learned ring aggregation model, which relies on sampling heuristics. The Leefbaarometer only applies spatial convolutions and decontextualises variables to create a model suitable for all of the Netherlands. While spatial context matters in residents' perception, the Leefbaarometer does not account for it to the same extent.

Additionally, preliminary experiments showed that it is possible to compress origin-destination matrices filled with travel times. It seems learnt aggregation creates embeddings similarly, much like step three of learning strategy two. It captures the relative position of spatial units in some subspaces of the urban representations. The relative position does not need geospatial coordinates but can be higher dimensional to capture, for example, travel times.

Of all modelling decisions, simply increasing the dimensionality shows the greatest return in predictive accuracy. Only when exceeding a few hundred dimensions will performance start to suffer due to the ratio of spatial units to dimensionality. Nevertheless, this thesis would have much higher results if the dimensionality used in the linear regression was not 30 but a multiple of it, such as 100 or even 250.

Over-Squashing

In learnt aggregation, we assign an equal number of parameters to each k-ring. As the number of spatial units increases linearly by a factor of six for every increase in rings, this might lead to over-squashing, similar to what is seen in graph convolutional neural networks. Oversquashing occurs when too much information is crammed into too few weights, particularly in message passing over longer contexts (Alon & Yahav, 2021).

Future work may explore fixing the number of parameters per ring rather than sharing these. Many advanced AI models apply similar reasoning; data representations are progressively quantised further away in time or space (Dettmers et al., 2021). For example, fewer resources are allocated to events forecasted far into the future or far away, as is done by Tesla Autopilot.

A way to tackle the problem of bottlenecks is to concatenate representations rather than aggregate them, preserving all of the information before feeding them into a very wide neural network. The initial implementation of ring aggregation followed this approach and performed reasonably well. Nevertheless, it is hard to experiment with this setup as the number of parameters changes drastically depending on the number of k-rings evaluated. Additionally, performance did not change much when the number of weights was alternating. Therefore, it may be tentatively stated that the generous application of concatenation is not a good way to address the bottleneck problem.

To avoid concatenating the results of spatial units, one could rely on multiple statistical moments to maintain a fixed number of parameters using a neural network for within-ring aggregation. Rather than relying on a single statistical moment within each ring—aggregating using the mean—it may be more effective to consider the first two or three statistical moments. The distribution of representations per ring is then described using the mean, variance, and skewness.

Accessibility

The lack of public transport modes (bus, tram, train, metro) significantly affects our urban representations. Large parts of Rotterdam and The Hague are more integrated with their respective cities, but our learnt aggregation representations do not reflect this.

Travel times for walking are not representative of reality, especially near Gorinchem, see Figure 86. In learning strategy one, we sample triplets using location-based accessibility, with the top two per cent being positive and the rest being negative, averaged across walking, cycling, and driving networks. The abnormally high levels of accessibility in this region lead to biased results in the bottom right corner of the study area. However, since it is on the edge of the map and isolated from the other cities, it was deemed ok not to investigate the cause for this error during the limited time allotted for the thesis. It is likely related to travel time calculation.

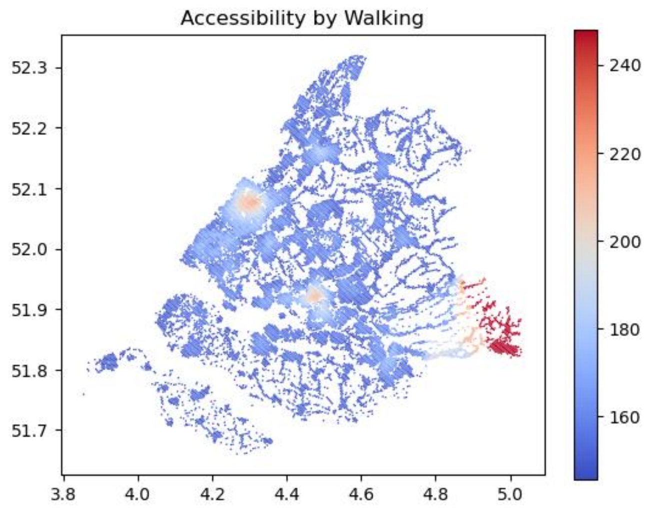


Figure 86: Calculated accessibility for the proximity measure used in learning strategy one. Note the abnormally high accessibility in the bottom right-hand corner (only a large town and countryside).

The results of learning strategy two showed that when using location-based accessibility, Zoetermeer was clustered with the ‘Groene Hart’, a more rural part of South Holland, Figure 84. This finding is in line with Zoetermeer's policy vision (Gemeente Zoetermeer, 2017), which notes the lack of connections with Rotterdam. Zoetermeer was originally a satellite suburb of The Hague and is, therefore, not as well connected to the remainder of the Randstad as it could be.

6.4 Implications for Transport Policy

The dynamic approach to liveability developed in this thesis has significant implications for transport policy and urban planning. It aligns with emerging concepts in governance and policy-making while presenting challenges for practical implementation.

The dynamic approach aligns with the concept of *'uitnodigend bestuur'* (inviting governance) in transport policy (Buuren, 2018). By attempting to model complex interactions within urban systems, it may be possible to develop more adaptable and participatory policies. Inviting governance differs from more traditional top-down planning approaches by explicitly considering *‘twee benigheid’*. This concept suggests that governance should not only fulfil its pragmatic duties, like upholding laws and administrative processes but also maintain openness towards new perspectives—a form of epistemic foraging. In this view, the amorphous complex system embodied by governance, inhabiting its niche, forages the world comprised of residents as a whole.

This shift from traditional top-down planning approaches is particularly significant when considering the underlying mechanisms of active inference. The predictive processing framework operationalises the dynamics of message passing in active inference. Predictive processing introduces additional dynamics that need to be considered in policy-making. Predictive processing is explicitly a two-way, considering both top-down and bottom-up processing. These dynamics of error propagation have already been studied for organisations by Khezri (2022), who refers to it as *‘governing continuous transformation’*, a concept applied to the Dutch case by Braams (2023).

The dynamic approach also relates to the concept of *'homo narrans'* in transport policy (Schwanen, 2020). This view considers residents as narrative beings who construct meaning through stories and experiences, contrasting with the *'homo economicus'* model of purely rational economic actors. 'Living' digital twins rely on the same narrative structuring. Active inference describes a narrative machine, whereas discrete choice models describe a rational machine. It is worth noting that this narrative-based approach aligns with how some policymakers already use cost-benefit analysis (CBA). Many view CBA as a tool to account for relevant factors rather than relying solely on monetary results (Mouter, 2017). The elements included in a CBA often indicate what is considered necessary in decision-making. Therefore, incorporating narratives more directly into modelling frameworks might align with political practice.

6.5 Future Directions: Towards Living Digital Twins

Complex systems modelling and searching for alternative intelligence collide with studying living digital twins. That is, modelling the niche construction of existing complex systems by deploying living digital twins. Such digital twins involve generative models as proposed under the active inference framework. Hipolito & Khanduja (2024) and Albarracin et al. (2024) propose tackling wicked problems like sustainability using digital twins of complex dynamical systems. Hipolito and Khanduja start by noting the limitations of reductionism, splitting the world into many variables (indicators) to be measured and related to each other. At the same time, Albarracin et al. approach the problem by proposing bounds on resources as target states.

On the other hand, there is the search for alternative intelligences; in the work "Technological Approach to Mind Everywhere: An Experimentally Grounded Framework for Understanding Diverse Bodies and Minds", Levin (2022) synthesis recent results in biology and applications of active inference, reframing the notion of multi-scale action-perception loops. Fields & Levin (2022) generalise behaviour to illustrate how intelligence is the navigation of some (abstract) search space, whether the journey of a resident using the transport network of a city or morphospace. Morphospace contains all physical bodies, explaining why the limb of a salamander grows back until exactly where it should stop. Somehow, cells and the body all have niches they want to inhabit. Its cumulative effect leads to the construction of a body until it is satisfactory.

There is a distinction between things that live, explicitly performing inference over their generative model and those that only follow free energy gradients to optimise their dynamics. As such, it is not the correct question whether living digital twins involve syncing their niche construction to actual living or inference-performing systems. Instead, all that matters for applying the active inference system to operationalise niche construction is whether these systems perform gradient descent over free energy, leaving the notion of living as a useful explanatory fiction rather than a statement about the true nature of the world (M. J. D. Ramstead et al., 2024). It does, however, evoke a more fundamental question. How and to what extent do spatio-temporal scales, as in multi-scale action-perception loops, influence each other? It is the study of the relationship between these niche construction systems, or things in the free energy principle, which will determine if it is sufficient to model the niche construction process of entire cities as a proxy for the liveability of its nested residents.

The action-perception loop is central to the dynamic approach. The action-perception loop describes the reciprocal relationship between residents and their living environment. Moreover, the interaction between residents and the environment is only one of many at a particular scale. Lower scales would involve the internal dynamics of the resident or environment. On the other hand, higher scales abstract the entire relationship away, such that the internal dynamics of an entire urban region are of interest. Using complex systems modelling, operationalised with active inference, to occupy the niche construction process of existent things could be viewed as discovering alternative intelligence. That is, living digital twins should only work if something else shares a proximate niche. Otherwise, one starts a new niche that has not existed before. However, solely occupying non-interfering niches may be sufficient for infrastructure engineering purposes. The definition of non-interfering niches is left to be studied later. Let alone that of positive or negative interference, memes and anti-memes.

There are no real-world implementations of urban living digital twins. The mechanism that these twins take advantage of is niche construction. However, to perform niche construction, one needs something with a transmission and needs/desires to get it going. The acquisition of these two components is a major obstacle in deploying living digital twins. First, the transmission is only really required for real-world systems since toy examples do not need neural networks. Instead, toy examples can work with indicators directly, which leaves needs/desires as the most challenging obstacle. The characteristic state of a niche aligns with the system's transmission; that is, the characteristic state cannot be identified without knowing which indicators and (initialised) likelihood mapping are available. A further complication is the addition of hierarchical models, which are well-suited for transportation modelling. Addressing these three obstacles will be the focus of the remainder of this chapter.

The dynamic approach to urban systems and transport policy offers a different perspective compared to traditional static methods. It invites the modeller to revisit the concept of seeing like a state (Scott, 1998), applied to active inference by Avel Guénin—Carlut (2022). Seeing like a state asks the modeller to consider what is measured (indicators) and its impact on development. More specifically, we would consider seeing the transport system. To temporarily believe that complex systems are niche-constructing things allows for the creation and engagement of a ‘living’ digital twin. To take the work of M. J. D. Ramstead et al. (2024) a step further, entertaining the useful explanatory fiction that minimises free energy across a set of parameters is as if making actual inferences.

A significant policy objective afforded by considering the transport system a niche constructor would be its simplification. Active inference formalises the calculation of variational free energy, a measure of divergence between expected and actual observations, which is composed of maximising predictive accuracy while minimising model complexity. When made explicit, it will invite modellers to simplify the transport system in addition to fitting the available data as accurately as possible. That is, the digital twin of the transport system attempts to be low complexity. When implemented, the transport system’s digital twin acts, self-evidencing for its world model. Thus, a digital twin that tries to become simpler may do so by making the actual transport system simpler through yet-to-be-defined avenues. However, a significant knowledge gap exists regarding whether and how this coupling occurs.

In total, we distinguish three primary approaches to living digital twins, each with its strengths and challenges:

1. Modelling individual residents: This agent-based approach provides granular insights into individual behaviour and decision-making. However, it is computationally intensive and requires high-fidelity digital twins to approximate real-world dynamics accurately.
2. Treating spatial units as agents: This approach aligns with the methodology developed in the current thesis, particularly due to local aggregation, which was developed in learning strategy one. Each hexagonal unit is modelled as an agent with needs and desires, observing and acting on its local environment. Agglomeration effects between spatial units increase cognitive lightcones, how far into space and time a decision maker can consider (Levin, 2019).

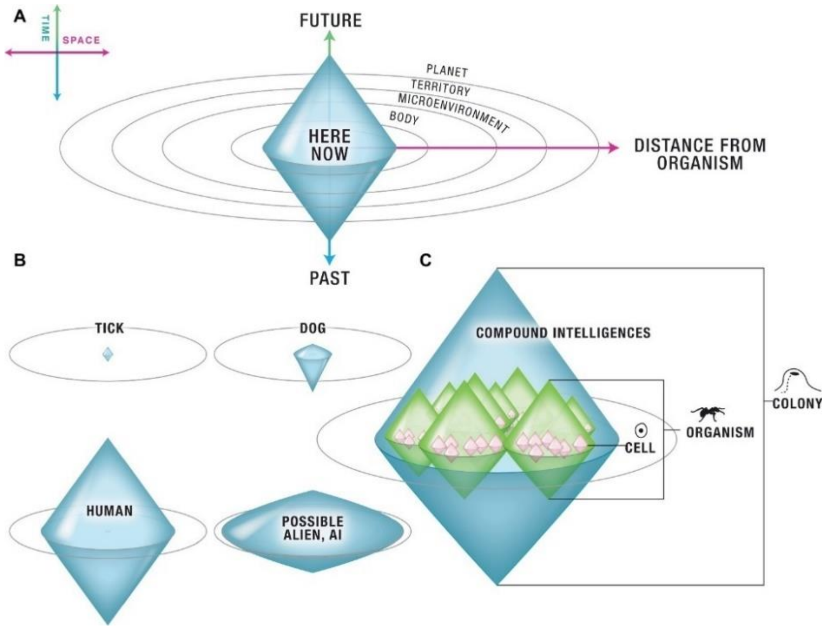


Figure 87: Computational boundary of self. From (Levin, 2019).

3. Rather than simulate the agglomeration of spatial units into a hierarchical organisation, it makes more sense to start with one, considering the entire urban region as a single generative model. It treats the city as a living, adapting entity. Different layers in the hierarchy correspond to different spatio-temporal scales. Since liveability concerns local scales of the action-perception loop and sustainability covers global scales, hierarchical active inference models may potentially address the mesoscale sustainability gap identified by Johnson et al. (2023).

6.5.1 Active Inference Models are Transport Models

In line with the third approach to model living digital twins, as outlined in the previous section, we now entertain the hypothesis that hierarchical generative models are implicitly transport models. The total assembly of such a model requires a transmission and engine. In line with current transport modelling practice, the 4-step transport model covers the generation, distribution, travel mode choice and assignment of travel volumes across the network. The first three steps up until travel mode choice are equivalent to location-based accessibility. The fourth step subsequently takes accessibility for the different travel modes per spatial unit and assigns the volume of journeys across the transportation network links.

This hypothesis arises from the observation that both active inference and transport systems fundamentally deal with the movement of representations, whether that be a vector as used in neural networks or the number of cars as used in transport modelling. Furthermore, modellers always have to define the boundary of the system which is being modelled. Hence, the notion of a Markov blanket separating internal and external states comes very naturally. It is only a matter of defining and operationalising this blanket's sensory and active states, which becomes problematic due to the vast knowledge gaps.

We propose implementing a hierarchical active inference model using the H3 geospatial index to test this hypothesis. Hexagonal spatial units of the highest manageable resolution are in contact with the world, serving as the interface between the model and reality. At the same time, the lowest resolution hexagons are first in line to receive messages loaded with preferences (needs/desires), the configuration of which should likely be aligned with broad prosperity indicators.

Active inference models may be understood as a sandwich with two pieces of bread that act and a filling which does bidirectional message passing. The bread in this model would be the highest and lowest resolution layers. More formally, it is a sandwich of screens (M. J. Ramstead et al., 2023). See also Figure 92. Each layer or screen is a landscape of affordances, from which actions are selected top-down and prepared using bottom-up feedback. The dynamics of attention are, however, more nuanced and based on the field of affordances, which is constrained to all which was relevant only recently. Thus, attention flows literally and does not jump around from one end of the landscape to another.

To understand the flow of attention and its analogy to transport modelling, we note that the hierarchical structure mirrors the small-world nature of self-organising Bayesian systems. Consider going up in the hierarchy, such as taking a highway onramp or accessing a train. Going down vice versa, e.g. moving from an arterial to a local street. Short trips do not need to go up and over, instead taking a shortcut as indicated by the grey lines in the side view of Figure 88.

Circling to *homo narrans* as it relates to transport policy offers a compelling contextualisation. A narrative is the flow of beliefs up and down throughout the hierarchical network. A travel journey as studied in person-based accessibility (K. T. Geurs & van Wee, 2004) becomes itself the object of interest in the paths-based formulation of Bayesian mechanics, where such journeys of attention are modelled as a whole, albeit probabilistically (M. J. D. Ramstead et al., 2023).

Internal network dynamics often follow power-law distributions (Goekoop & de Kleijn, 2023), as seen in accessibility (Geurs, 2018). Furthermore, hierarchical Bayesian control systems may collapse during long exposure to stressful events, leading to functional disintegration (Goekoop & De Kleijn, 2021). The 15-minute city concept and re-/friendshoring of supply chains can be seen as a means to lower stress on the system. Dynamic transport models could, therefore, offer novel techniques to address resilience from a transport engineering perspective and a policy one through sustainability, as resilience is one component of sustainability (Albarracin et al., 2024). The inclusion of hierarchy makes this proposed approach distinct from previous efforts to simulate traffic systems using generative models (Wong & Farooq, 2019).

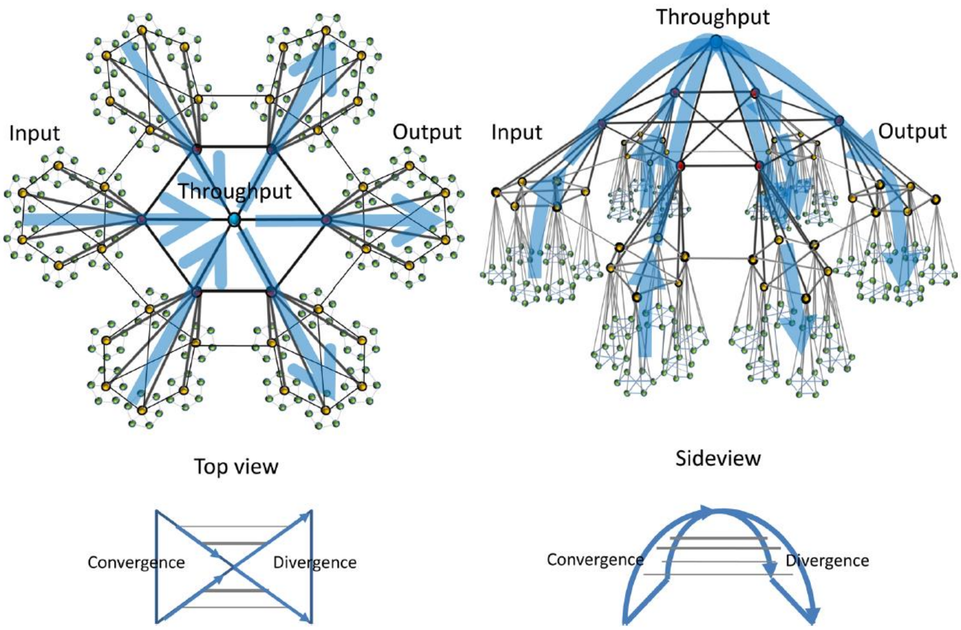


Figure 88: Small world network and internal dynamics of hierarchical Bayesian control system. From (Goekoop & De Kleijn, 2021).

The implementation of such a hierarchical generative model can draw upon recent work by K. Friston et al. (2024), who developed a renormalising generative model with discrete states (Figure 89). The key point of relevance is that both the H3 geospatial index and renormalising generative models have parent-child relationships between resolution layers, such that one has to go up and over to visit a child from another parent. The group or block in the dashed red box represents the children of the blue circle above it. Each spatial unit (hexagon) will learn unique weights towards its parent and children, the D matrix. If these weights were not unique, then the model is homologous to (deep) convolutional neural networks—as was studied in this thesis.

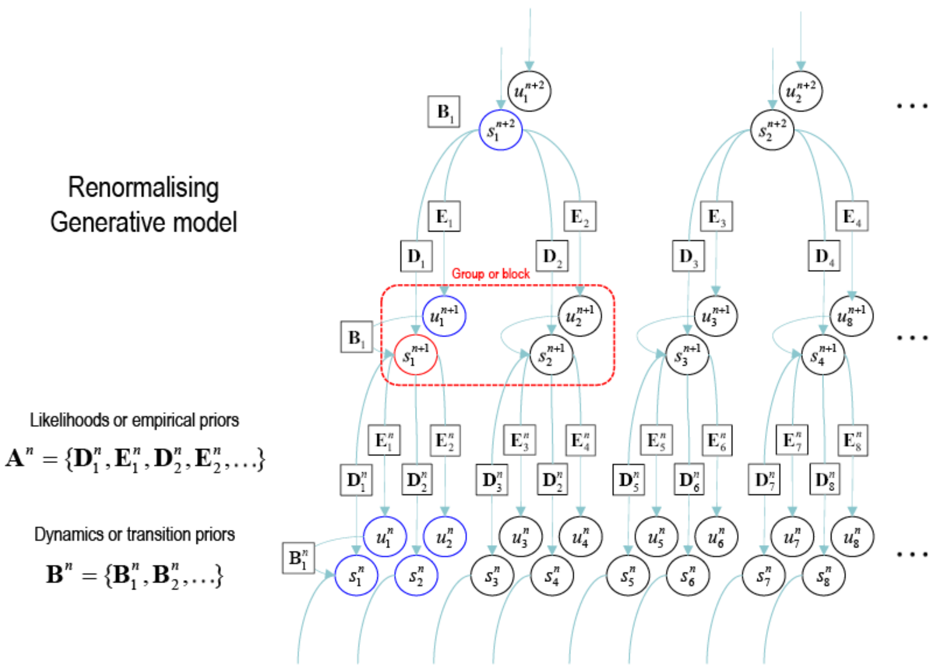


Figure 89: Illustration of a renormalising generative model capable of mapping pixels to actions. From (K. Friston et al., 2024)

6.5.2 Total Assembly

Operationalising the dynamic approach to liveability requires the construction of an assembly comprised of transmission and engine. The previous section on transport modelling argued that the transmission captures the first three steps of the 4-step transport model. The engine fulfils the final step, which is assignment. This section details the total assembly, explaining the purpose and interaction of each component. Subsequent sections on transmission and engine provide detailed descriptions of their configuration and training.

First, the dynamic approach to liveability considers perception itself an action. The fit between residents and the environment is placed at perception, between indicators and percepts. The influence of needs/desires is subsequently one of bias, a bias applied to percepts to align with the characteristic states of the niche. The purpose of the transmission is, therefore, to be in contact with the world dynamically. Meanwhile, the engine traverses the landscape of affordances provided by the transmission.

Several steps are overlooked in this treatment of the assembly. First, the transmission will be much more detailed than studied in this thesis, requiring a unique workflow per hierarchical layer (H3 resolution). Second, the engine requiring action selection is far beyond the scope of this thesis. See also (Browne et al., 2012; K. Friston et al., 2021; Koudahl et al., 2023).

The transmission must be able to capture low-dimensional regularities in high-dimensional data, either through the construction of indicators using feature extractors where the ambiguity of mappings is low or by mapping indicators to percepts where ambiguity is high. Both indicator and percept can be representations, as has been the case in this thesis under learning strategy one. The fusion network is shown as the neural network block labelled perception and is to be updated through a modified variational free energy term (VFE), which optimises for accuracy and complexity. Alternatively, the selection of actions from needs/desires is determined by expected free energy (EFE). Actions constrain the landscape of affordances, which is as flat as possible to allow for option value, as studied in transportation planning (K. T. Geurs et al., 2006).

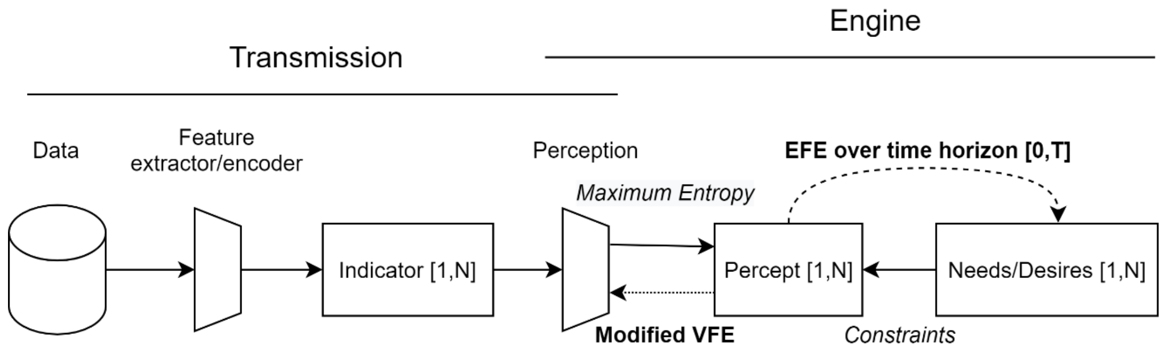


Figure 90: Information engine for the dynamic operationalisation of liveability using transport modelling.

6.5.3 Transmission

The transmission exists to map high-dimensional data into manageable percepts. It is comprised of two components, both capable of accommodating neural networks. Indicators benefit from low ambiguity mappings using the first encoder block. On the other hand, the likelihood mapping, where the transmission and engine interface should be left to address high ambiguity mappings. Otherwise, there would be no epistemic value to capture as part of EFE.

On the whole, the transmission should approximate the ring aggregation methodology developed as learning strategy one in this thesis. It comprises two steps. The first is learning representations for each data modality of the urban region, to be reframed as indicators. The second is the combination of these views into percepts and accounting for location-based accessibility in doing so, approximating the constrained maximum entropy principle.

Indicators

Indicators can be inferred by extracting features from high-dimensional data. Not only is it possible to use self-supervised techniques as in this thesis, but If the objective is to create interpretable indicators, they could be trained using a supervised loss with an existing indicator as a target. An image encoder network like ConvNeXt (Liu et al., 2022) can be trained to extract features from aerial images correlated with the Leefbaarometer's housing stock score. While our study's performance was limited, previous supervised learning approaches have shown success (Levering et al., 2023).

Alternatively, it is possible to use supervised learning techniques such as classification and segmentation to extract features. As such, indicators can automatically be generated from street view images to count the number of cars or people (Garrido-Valenzuela et al., 2023). Alternatively, the presence of urban canyons can be assessed by considering the frequency and intensity of greenery and sky (Gong et al., 2018).

Fusion into Percepts

Approximating the likelihood mapping of generative models can be done using a fusion network. Likelihood mapping is a matrix which relates observations to representations. The mapping should be biased towards affordance landscapes using location-based accessibility. All that matters is that the representation fed into the active inference model covers the choice set as accurately as possible. In turn, the hierarchical active inference model goes up and over to assign travel volumes between spatial units, see Figure 88.

The calculation of location-based accessibility involves resistances between spatial units and the attractiveness thereof. Both of these deserve extensive attention in later work. First, resistances between spatial units have been operationalised using travel time in this thesis. Alternatively, one could use generalised travel costs or another measure of travel resistance that is still to be imagined. Second, attractiveness can be more than just the number of jobs, shops or building density. Any form of utility assigned to a spatial unit could be used—for example, visual quality (van Cranenburgh & Garrido-Valenzuela, 2023).

Sampling Heuristics

The sampling heuristic can be significantly improved to gather positive and negative pairs for similarity loss used in the fusion network. The methodology tested in this thesis used both 2-step ring sampling and weighted random walks to operationalise location-based accessibility. 2-step ring sampling to finetune the aerial image encoder network (ConvNeXt model). Weighted random walks across increments of location-based accessibility in learning strategy two. The modelling study performed in this thesis found empirically that the highest predictive accuracy for liveability was achieved with finetuned aerial image embeddings. Learning strategy one relied on a cutoff value of the top 2 percentile accessible spatial units. The results of learning strategy two indicate similar results to those of learning strategy one.

Since the transmission is concerned with crafting landscapes of affordances, whatever H3 resolution of interest, we should fully take advantage of the hexagonal grid to implement location-based accessibility in a future-proof manner, fit for introducing novel, more expressive forms of travel resistance and attractiveness. To do so, we observe that location-based accessibility is concerned with increments of both attractiveness and travel resistance. The more increments accumulated across the attractiveness space, the more likely a trip is. While the more increments of travel resistance accumulate across space, the less likely a trip is. Hence, location-based accessibility is somewhat like a voltage, which drives flows/currents in the active inference model as attention flows between the hierarchical layers, top-down/bottom-up. See also Parvizi-Wayne et al. (2023) for the phenomenology of flow and active inference. The transmissions, one for each, prepare a field of affordances for every hierarchical layer, and the engine calculates paths between the hierarchical layers.

We propose to continue off of the 2-step ring sampling methodology by fixing a receptive field. However, future work may show that the size of the field itself can be dynamic, too. The field spans a set of spatial units, which are doubly assigned to two weighted graphs. One graph has increments of attractiveness on its edges, whereas the other has increments of travel resistance. Random walks are then performed to obtain the result of 2-step ring sampling in the case of a perfectly uniform graph. Random weighted walks for the sampling of negative triplets start outside and go inwards. Whereas random weighted walks for the sampling of positive triplets start on the inside and go outwards.

At the end, not shown, the two plots are combined to classify each spatial unit for a given centre. Positive and negative sampling occurs, and a threshold is applied to classify the final sets to be used in triplet sampling. That means an equally often sampled spatial unit will be left neutral. Spatial units in the receptive field of the blue centre are either neutral, to be positively sampled or negatively sampled. Different numbers of walks may be sampled for either. After the classification, triplet mining makes it possible to obtain the most difficult spatial units in the context and improve the fusion network's training process.

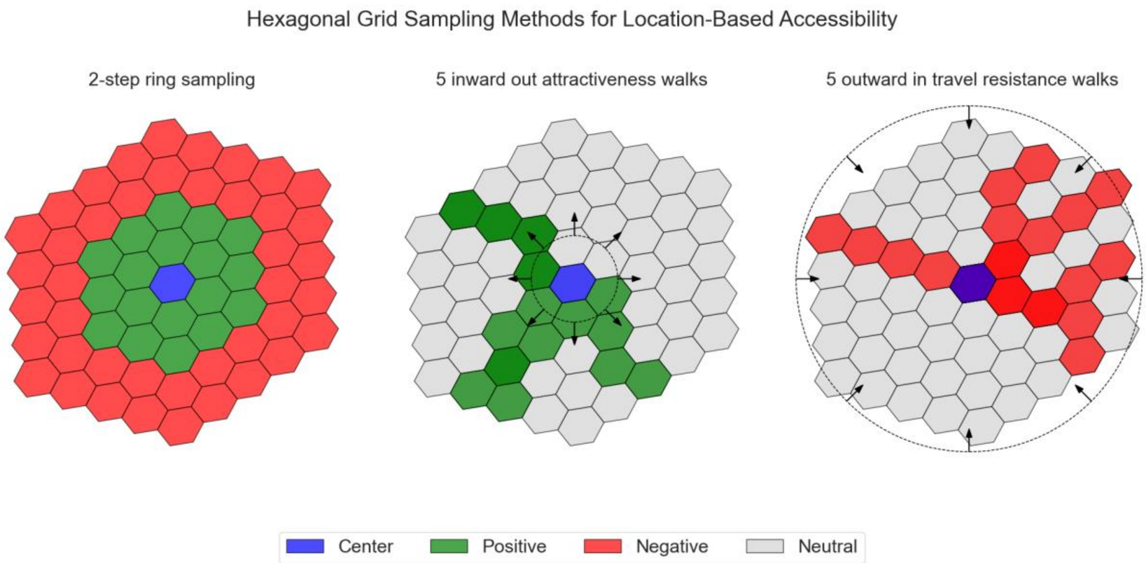


Figure 91: hexagonal grid sampling methods for location-based accessibility. Weighted random walks are based on attractiveness inside out and travel resistance outside in.

Expressivity

Architectural improvements are expected to improve expressivity when it comes to crafting representations. Transportation involves discrete events and objects—for example, the rate of travel or pollution throughout space over time. Rate coding has been applied to representation learning by implementing the Poisson distribution, which is commonly applied to queuing problems. Rate coding has already been used in this thesis using the Geovex encoder for points of interest by Donghi & Morvan (2023). Active inference models optimise statistical moments of probability distributions, which parameterise states. In turn, it is sensible to consider variational autoencoders optimising for the mean and variance of Gaussian distributions. Poisson Variational Autoencoders by Vafaii et al. (2024) offer a means to capture discrete transport phenomena probabilistically, effectively modelling the intensity of presence or absence in urban spaces.

6.5.4 Engine

At the intersection of niche construction and transport policy is the notion of information engines and their associated transmissions. Active inference models, which build upon the free energy principle, retrieve a great deal of physics from thermodynamics and statistical mechanics, both the foundations of conventional combustion engines. For now, it should suffice to claim that such dynamic models are information engines that perform work in the sense that they construct and maintain their niche, given a set of indicators and needs/preferences. This thesis attempted to understand the nature of these transmissions: neural networks whose purpose is to map high-dimensional data to representations for input to an active inference model.

The construction and maintenance of a niche is abstract but clarified by considering that agents which share a set of affordances share a niche. That is, they both have similar capabilities for action and, in turn, influence the world the other inhabits and perceives. Such information engines take in energy to run the computers and return free energy gradients of a relevant niche. The niche, however, has to be retrieved by carefully selecting the indicators and needs/desires. This last bit is neither simple nor evident, and much work remains.

The transmission (indicators) and needs/desires belonging to a niche should be inferred from revealed behaviour. Given that we are interested in hierarchical models, the obstacle is to reduce the degrees of freedom in the model, leaving only the transmission and needs/desires to be configured. To do so, living urban digital twins may benefit from continuing Spinoza's arguments. First, he introduces a single substance from which two perceivable attributes come forth: extension and thought, which mirror the duality of the free energy principle with external and internal states, respectively. Second, Training hierarchical active inference models of urban regions may significantly benefit from the complimentary duality of *natura naturata* and *natura naturans*. Each layer in the hierarchy contains the outcomes of the dynamic process, representations which parameterise beliefs about states. Supervised learning of internal representations could be possible by crafting a suitable transmission for each H3 resolution. These targets approximate the landscape of affordances for every level as they are trained using the newly proposed sampling procedure in Figure 91. We remove the notion of representations altogether by putting every level in the hierarchical model in contact with a transmission rather than just the bottom. The generative model is then fully dynamic and radically enactivist, with no representations, just feedback loops. See also section 2.5.4 on Integrating Predictive Processing and Enactivism for background.

Hence, representations obtained using static liveability are used to bootstrap the operationalisation of the dynamic approach to liveability. Supervised learning may improve our ability to find the characteristic states and indicators to approximate a niche of interest. Once the characteristic states and indicators are set, a good likelihood mapping is needed to create percepts, the second part of the transmission. Then, modelling can resume without inference on intervening hierarchical layers but only on the bottom (highest resolution H3). In practice, it may require gradually removing supervised internal layers, potentially even partially removing them within each layer.

After the trained model with configured transmission and needs/desires has run for a bit, one can extract the representations of the intervening layers and observe them as static representations of liveability. With a valuation according to the spatio-temporal scale befitting of it. The correspondence of valuations to spatio-temporal scales is expected since nested action-perception loops provide top-down constraints and bottom-up sensory data. That is, top-down action selection uses affordances, constraining what is perceived using the observations. It is the process of turning the generative model into a recognition model, a tale of two densities, see (M. J. D. Ramstead et al., 2020). Affordances are the choice set, and alternatives within them are evaluated for their attributes (representations) based on their pragmatic and epistemic value—where pragmatic value is most proximate to valuation as understood in the static approach to liveability. The pragmatic value concerns the characteristic state of the niche, preferences and needs/desires.

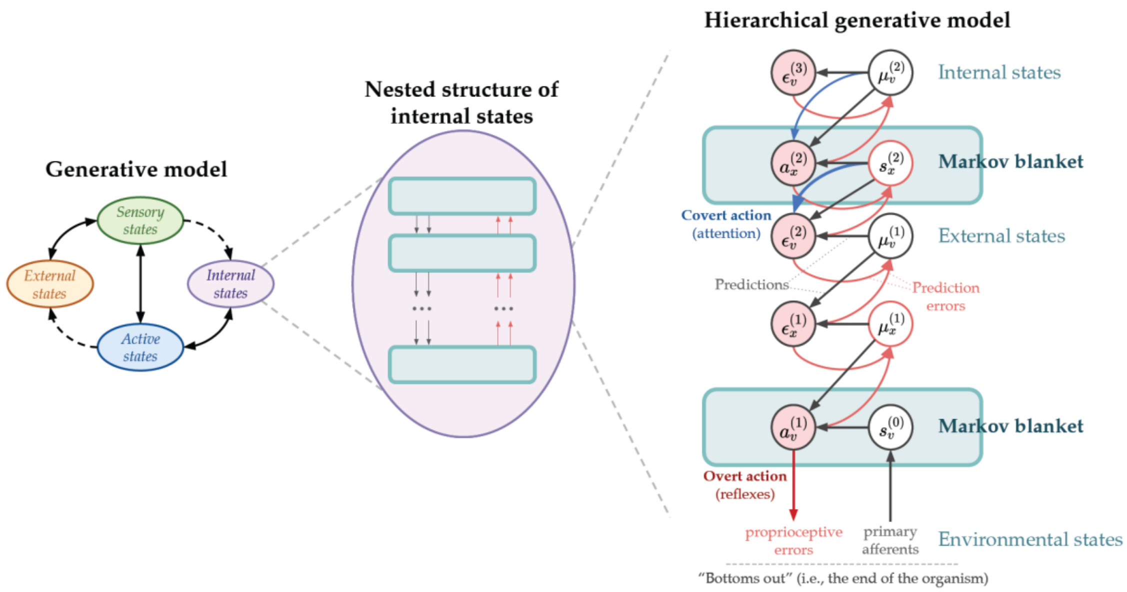


Figure 92: Inner screen theory of consciousness applied to active inference. Layers in the hierarchy are screens that combine top-down and bottom-up interactions. The highest layer only acts top-down (mental action), setting the characteristic states. The bottom layer is the only bottom-up, providing contact with the environment (physical action). From (M. J. Ramstead et al., 2023).

6.5.5 Final Considerations

The development of such models presents significant challenges, including data integration across multiple sources and modalities, computational requirements, and the need for novel validation approaches. Moreover, ethical considerations surrounding privacy, equity, and the potential unintended consequences of AI-driven urban planning must be carefully addressed.

Hijacking a niche requires syncing with it by identifying suitable indicators and needs/desires. In hierarchical models, learning the static representations of different hierarchical levels is another knowledge gap. Somehow, it is up to transport modellers to capture the spatiotemporal characteristics of each hierarchical layer by taking inventory of all relevant urban phenomena and collecting data continuously as necessary for the initial inference period of the active inference model. Lastly, the correct transmission architecture should be figured out to represent this data so that it fits with the other layers.

Practical

Generative models have come a long way. The newest developments suit engineering applications adopting reactive message passing (Koudahl et al., 2023; van de Laar et al., 2023). Reactive message passing is much cheaper since recurrent functions do not have a fixed timeline but can arrive anytime. E.g. a discrete-event model used in operations research has pre-determined timesteps 1, 2, 3,... but a reactive program does not. The continued development of reactive message passing in Eindhoven is highly valuable. Unlike traditional neural networks, active inference models provide an explainable framework crucial for transparency in decision-making processes (Albarracin, et al., 2023).

Cognitive Security

Living digital twinning represents a significant advancement in complex systems modelling involving hijacking niche construction processes through artificial entities. This expansion of system surfaces towards niche-embodying elements raises questions about cognitive security and the potential impacts of interference with these digital constructs.

The concept of living digital twins extends beyond traditional physical infrastructure, encompassing high-dimensional representations of travel impedances and attractiveness. Travel impedances may incorporate complex resistances between spatial units that go beyond mere travel time or generalized cost, potentially involving high-dimensional metrics. Similarly, attractiveness might be conceptualized in a 1000-dimensional metric space rather than relying solely on conventional measures like job numbers or building density. This sophisticated approach to modelling accessibility enables a more nuanced understanding of how individuals interact with their environment.

Such advanced modelling capabilities facilitate the parametric design of infrastructure and built environments, albeit with inherent biases towards stakeholder-defined characteristic states. The automation inherent in this approach necessitates a critical examination of its objectives and search space navigation. The distribution of affordances—derived from these complex accessibility measures—across spatial, temporal, and demographic dimensions may significantly influence resultant behaviours.

For instance, the classic example of transport justice—where a low-hanging viaduct prevents bus access to a beach, disproportionately affecting low-income residents—illustrates how physical constraints can serve as initial modelling parameters for affordances.

However, the ability to engineer affordances in high-dimensional spaces introduces a form of nudging far more subtle than explicit physical barriers. This nuanced approach to behavioural influence raises concerns about privacy. Perception is action such that to tamper with the landscape/field of affordances is to remove privacy before any action has occurred yet.

Municipalities in the Netherlands (VNG) have voiced concerns with this type of autonomous digital twin (Vereniging van Nederlandse Gemeenten, 2024). VNG recommends the study of frameworks and guidelines in the development of such models to understand what values and oversight are involved. Hipólito & Podosky (2024) propose a comprehensive overview of value and control in artificial intelligence. Indeed, autonomous living digital twins may take actions nobody understands, for which no one is held accountable. If perception itself is an action, then the entire model and its data stream should be continuously monitored for accountability.

On the upside, however, there may be the potential to create a new kind of polder in a more abstract space of affordances/capabilities in which niches are synced to and nudged towards global optima. In this case, the global optima is set by larger information processing hierarchies such as an urban region, entire country or supranational network, as proposed by the spatial web using ecosystems of intelligence (K. J. Friston et al., 2022).

Polder may be a suitable and hopeful reference since it is a practice of pumping free energy gradients to improve liveability; however, the practice may be extractive with considerable externalities. Niches are more than the relationship of residents with their urban environment. Niches are ubiquitous as they describe the reciprocal relationship between any sense maker and their environment. It is unknown what occurs when artificial things mess with the niche construction process of existent things by hijacking their niche, potentially without consent. At this point, the development of cognitive security comes into play, which is the study of niche construction and the defence from interference by third parties. It takes inventory of ways people and intelligent systems can be nudged towards specific belief systems or brought to be unbalanced (Waltzman, 2017). Practical recommendations for (niche constructing) digital twins involve unified data infrastructure, professionalisation of this upcoming field, and securing sensor systems, as the latter is the most vulnerable to interference (Cordes, 2024).

Distributive Justice

The broad prosperity framework explicitly calls for distributive justice (de Boer et al., 2023; Raad voor de leefomgeving en infrastructuur, 2024; Snellen & Bastiaanssen, 2021). Distributive justice from a broad prosperity perspective starts with a critique of utilitarianism, which underpins current static approaches based on valuation. Aggregate satisfaction improvement in a population without accounting for the distribution of improvement leads to a winner-take-all effect. In the past, for example, infrastructure investments have mostly gone towards the Randstad instead of the provinces due to higher returns on investment.

In the dynamic approach to liveability, where valuation has been replaced by perception as a means of fit, these concerns do not apply anymore. Instead, distributive justice becomes an issue of defining work. The work produced by niche-constructing information engines has to be quantified. The trouble here is that work is the expenditure of useful energy, but usefulness is subjective. Any concentration of capital and power will be able to ammas computational resources and sensory/actuator capabilities. Therefore, the dynamic approach to liveability does not inherently address the distributive justice component of broad prosperity; instead, it shuffles it around to whoever owns and configures the means of niche construction.

6.6 Synthesis and Future Prospects

This discussion has explored the complementarity of static and dynamic approaches to liveability, highlighting the potential of living digital twins in urban and transport systems modelling. The journey from urban representation learning to the conceptualisation of hierarchical active inference models as transport models reveals a promising path forward in operationalising liveability dynamically.

Key contributions include:

- The development of a theoretical framework bridging and contrasting the static and dynamic approaches to liveability.
- Empirical insights from urban representation learning, informing future dynamic operationalisations.
- The proposal of hierarchical active inference models as a novel approach to transport modelling.

Future work should focus on addressing the challenges in implementing living digital twins, particularly in developing appropriate transmissions, identifying needs/desires, and constructing hierarchical models suited for transportation modelling. Integrating these concepts with existing transport policy frameworks, such as broad prosperity, presents opportunities and ethical considerations that warrant careful examination. As with liveability, theory and practice influence each other, and one can cause lock-in of the other. The potential for living digital twins to offer new insights into urban dynamics, resilience, and sustainability remains a compelling direction for future research and policy development.

7 Appendix A: Scientific Research Paper

Learning Urban Representations to Operationalise Liveability

Bert Berkers

Abstract

Urban liveability is crucial in transport policy evaluation, but its operationalisation remains labour-intensive. This study introduces a novel urban representation learning method to automate liveability assessment, leveraging spatial convolutions on the H3 discrete global grid system (DGGS). Our approach uses deep neural networks and diverse data sources, including aerial and street view images, road network and public transport characteristics, and points of interest, to predict Leefbaarometer scores, a Dutch liveability assessment tool. We employ a late-fusion strategy, training the fusion network with circle loss and sampling triplets using Euclidean distance and accessibility heuristics. The introduced ring aggregation method outperforms existing approaches like urban2vec across all Leefbaarometer scores, with notable improvements in R-squared scores for amenities (0.85 vs 0.64), social cohesion (0.6 vs 0.43), and overall liveability (0.35 vs 0.27). Sensitivity analysis reveals that different Leefbaarometer scores perform optimally at varying receptive fields, and data source selection significantly impacts predictive performance. This research contributes to more efficient, automated liveability assessment methods, paving the way for dynamic, process-oriented models in urban planning and policy evaluation.

7.1 Introduction

Liveability has emerged as a crucial factor in Dutch transport policy evaluation, now considered alongside traditional objectives such as accessibility and safety (Huibregtse, 2021). Defined as the fit between residents and their living environment (Dorst, 2005), liveability presents significant challenges in its operationalisation for policy evaluation.

Collecting and processing indicators that characterise the environment is labour-intensive. Moreover, selecting relevant indicators requires strong theoretical backing to ensure that the measured aspects of the environment are indeed related to liveability. This complexity is further compounded by the need to obtain valuations of these indicators, although the automation of valuation falls outside the scope of the present study and is left to the recommendations.

The Leefbaarometer (Mandemakers et al., 2021) is the state-of-the-art operationalisation of liveability in the Netherlands. While valuable, the increasing pressure on urban transportation systems, driven by densification and the need for sustainable development (Gupta et al., 2024), necessitates more efficient and effective policy evaluation methods.

This research addresses this need by exploring the potential of urban representation learning to automate the operationalisation of liveability. We introduce ring aggregation, a novel urban representation learning method based on spatial convolutions using the H3 hexagonal DGGS. Our approach leverages deep neural networks and diverse data sources to learn compressed representations of urban environments and predict Leefbaarometer scores.

Research objective:

To automate the operationalisation of liveability.

Research questions:

1) How can urban representation learning be leveraged to automate the operationalisation of liveability?

2) What configuration of the proposed ring aggregation method most effectively predicts Leefbaarometer scores?

Through sensitivity analysis, we investigate the impact of various modelling decisions, including the choice of sampling heuristics, the configuration of spatial convolutions, and the selection of data sources. We also compare our method to existing approaches in the field.

By automating the creation and perception of indicators, we contribute to more efficient and data-driven policy evaluation. Furthermore, this research lays the groundwork for future explorations into dynamic, process-oriented models that could offer a more comprehensive understanding of liveability in urban environments.

The remainder of this paper is structured as follows: Section 2 presents the theoretical framework, introducing key concepts in liveability and urban representation learning. Section 3 details our methodology, including the novel ring aggregation method and our approach to multi-modal data fusion. Section 4 describes the application, including the study area, data sources and training procedure. Section 5 presents our results, analysing the impact of various modelling decisions and comparing our method to existing approaches. Finally, Section 6 discusses the implications of our findings and provides directions for future research.

7.2 Theoretical framework

Liveability is a complex and multi-dimensional concept encompassing various aspects of urban life. It is often used interchangeably with terms like quality of life, well-being, environmental quality, and health, leading to ambiguity in its definition. Quality of life and liveability are commonly used interchangeably (Tan et al., 2024). Both these terms aim to capture the extent to which an environment meets the needs of its residents. As De Haan et al. (2014) put it, *"Fulfilling human needs, and fulfilling more of them, increases the quality of life"*.

Leidelmeijer (2004) provides a succinct demarcation of terminology, starting with the foundational definition that liveability is the fit between a resident and their living environment. Taking a perspectival measurement from this reciprocal system yields liveability or quality of life, as shown in Figure 93. Sustainability and liveability are related but distinct in their scales. Both concern meeting needs (De Haan et al., 2014).

However, sustainability concerns a much larger spatiotemporal vista than liveability. The latter is reserved for local neighbourhood-scale relations between residents and their daily local environment—the here and now.

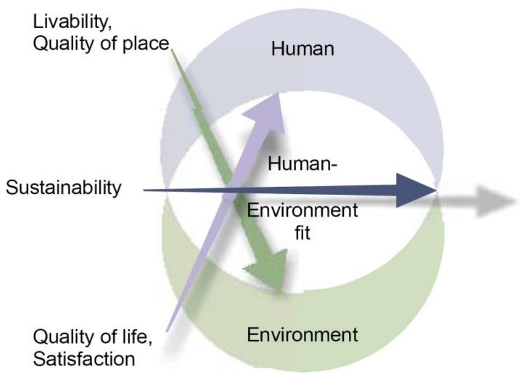


Figure 93: Reciprocal relationship between human/resident and environment. Perspectival measurement. From (Leidelmeijer, 2004).

Dorst (2005) outlined three operationalisations focusing on the role of indicators and residents' valuation of the living environment: perceived liveability, indicated liveability, and apparent liveability. 1) Perceived liveability relies solely on residents' valuations, which may be stated or revealed. Stated preferences are acquired through surveys or interviews, while revealed preferences are patterns found in residents' collective behaviour, for example, travel behaviour or house prices. 2) Indicated liveability, on the other hand, relies on a normative judgment by the analyst to determine what would likely make for a more liveable urban environment. Many operationalisations rely on this form in practice due to the ease of development. 3) Lastly, apparent liveability 'emerges' from the interaction between indicators and valuation. For example, it can be done by estimating a regression model between house prices and indicators. The Leefbaarometer by Mandemakers et al. (2021) is an example of apparent liveability; it combines indicators bundled into aspects and valuation through hedonic pricing and surveys on satisfaction with their living environment. The included aspects are safety, social cohesion, physical environment, housing stock, and amenities.

7.2.1 Conceptual Model

Liveability is inherently subjective. Which environment is conducive to fit differs per individual and throughout their life (Veenhoven, 2000). We propose the conceptual model in Figure 94, with three facets of subjectivity. First, indicators describe the environment. Having a combination of subjective and objective indicators is good practice. As Pacione (2003) states, *'we must consider both the city on the ground and the city in the mind'*. The leefbaarometer uses survey results for subjective indicators, such as the feeling of safety. Second, the perception of the environment distinguishes liveability from environmental quality, which is deemed objective (Leidelmeijer, 2004). Last, to characterise the resident in the resident environment fit, needs/desires exist to be satisfied. Needs relate to basic requirements such as food and shelter. Desires to culturally embedded and personally relevant wishes (Veenhoven, 2000)

However, the three approaches to operationalise liveability proposed by Dorst (2005) do not make perception explicit; instead, they solely rely on indicators and their valuation, see Figure 94. Admittedly, the implicitness of perception is valid for similar contexts and populations where all would understand indicators the same.

The resource-intensive nature of collecting indicators for comprehensive models has spurred the development of automated approaches. Aerial imagery, in combination with remote sensing, has been employed to predict Leefbaarometer scores (Levering, Marcos, Van Vliet, et al., 2023). This approach shows promise in predicting aspects related to physical characteristics and housing stock, though it performs less well for features recognisable by proxy, particularly amenities.

Beyond data processing, deep learning models also serve to more closely approximate human perception (Dubey et al., 2016). Recent advances have focused on using street-view images to represent the urban environment (Fan et al., 2023; Huang et al., 2021; Xiao et al., 2021; M. Zhang et al., 2020). Others used street-view images to operationalise the utility people experience in house relocation where utility is valuation, a proxy for the resident environment fit (van Cranenburgh & Garrido-Valenzuela, 2023).

Our proposed conceptual model builds upon existing frameworks by explicitly incorporating the role of perception in liveability assessment (Figure 95). By introducing neural networks to map indicators to percepts, we acknowledge the subjective nature of how urban environments are experienced. However, this model assumes a degree of uniformity in perception across the population and needs/desires, which may not always hold in diverse urban contexts. Future work could explore incorporating individual or group-level variations in perceptual processes.

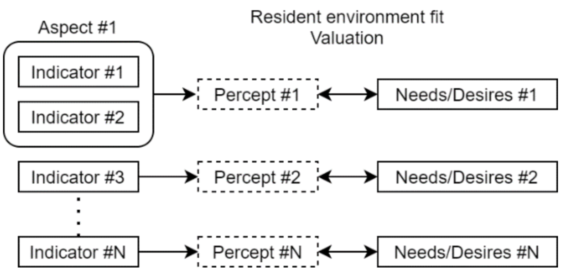


Figure 94: Conceptual model of operationalising liveability with valuation as a proxy for fit. Percepts are implicit, assuming uniformity across the population.

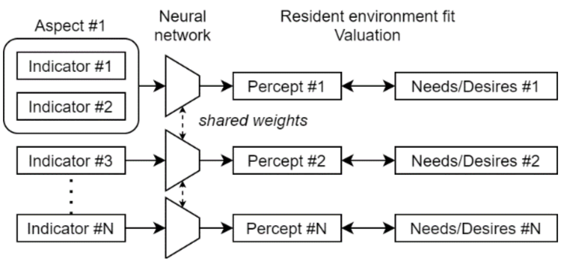


Figure 95: Conceptual model of operationalising liveability introducing neural networks to map indicators to percepts.

7.2.2 Learning Urban Representations

The field of urban representation learning employs machine learning techniques to generate compressed, computer-readable representations of complex urban environments. Urban representation learning is the study of mapping high-dimensional urban data into lower-dimensional embeddings while preserving the essential spatial and semantic correlations within the data. Examples of high-dimensional data used previously are street view images (Z. Wang et al., 2020), aerial imagery (Levering et al., 2023), points of interest (Donghi & Morvan, 2023; Woźniak & Szymański, 2021), the road network (Leśniara & Szymański, 2022) and the public transport network (Gramacki et al., 2021).

The first step in applying urban representation learning is selecting an appropriate spatial unit. Examples are administrative units like neighbourhoods or a discrete global grid system (DGGS). The H3 DGGS by Uber, with its hexagonal format and hierarchical structure, has gained popularity due to its ability to capture spatial relationships effectively. Furthermore, several studies use graphs to structure their data, as these can effectively keep track of relationships between spatial units to capture distances or travel volumes (Huang et al., 2021; Kim & Yoon, 2022; Xiao et al., 2021; M. Zhang et al., 2020). Graphs are also suitable for random-walk-based methods like node2vec (Grover & Leskovec, 2016) or graph convolutions (Kipf & Welling, 2017) and graph attention (Veličković et al., 2018).

The second choice is between supervised or self-supervised learning. The former is used in end-to-end applications when the task of interest is known. On the other hand, self-supervised learning is applied when a multi-step process is desired: representation of data and a subsequent downstream application. Supervised learning requires a labelled dataset with targets defined by the task, whereas self-supervised learning only requires labels distinguishing points within the data, hence the need for discrete spatial units.

Multi-modal learning enables the integration of information from diverse data sources, leading to more comprehensive representations which account for correlations between data sources and spatial units. However, it also poses challenges in effectively fusing different modalities. Various fusion strategies have been developed to address this challenge, including data-centric (Moschella et al., 2023), learned fusion approaches (F. Sun et al., 2023), and simple vector operations on representations like taking their mean (Raczycki, 2021). There is early and late fusion. Early fusion combines data sources immediately, whereas late fusion extracts features for each before combining them.

7.3 Methods

Building upon our conceptual model of liveability, which emphasises the role of indicators, perception, and needs/desires, we propose a novel urban representation learning methodology. This approach aims to automate the creation and perception of indicators through deep learning techniques. Our method applies late fusion. First, features that act as indicators derived from various data sources are created, and then, these are fused using the ring aggregation network to capture the spatial context of urban environments. This approach allows us to bridge the gap between traditional liveability assessment methods and the potential of urban representation learning. It builds upon the conceptual models developed in the theoretical framework by operationalising the automatisation of perception. As shown in Figure 4, the modelling pipeline expands on the conceptual model in Figure 95.

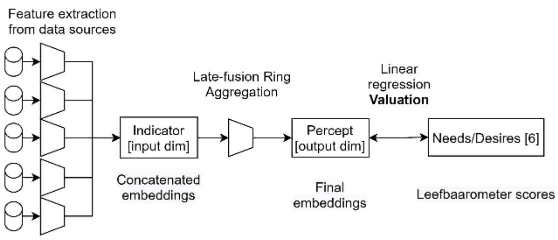


Figure 96: Conceptual overview of the modelling pipeline. Data flows from left to right until applied in a linear regression with Leefbaarometer scores.

7.3.1 Ring Aggregation

Our study introduces a novel urban representation learning methodology using H3 hexagonal spatial units called "ring aggregation." This method is inspired by the work of Raczycki (2021) and introduces learnable weights by mirroring graph convolutional neural networks but eliminates the need for explicit graph construction by utilising the neighbourhood indexing provided by the H3 geospatial index.

The advantage of ring aggregation lies in its ability to capture local context through spatial convolutions. Unlike traditional graph-based methods that rely on potentially irregular adjacency matrices, the H3 geospatial index provides a regular isotropic structure. This isotropy means that for any randomly selected hexagon, the surrounding rings are uniform in distance and number, allowing for a more consistent and interpretable aggregation of spatial information.

While traditional convolutions in image processing typically apply a learnable kernel to a regular grid, our ring aggregation method adapts this concept to the hexagonal structure of the H3 index. The primary differences are threefold. 1) Hexagonal grid: Instead of a square grid, we work with hexagonal cells, which provide better isotropy and more uniform neighbour relations. 2) Varying receptive field: Our method allows for easy adjustment of the number of rings considered, effectively changing the receptive field of the convolution. 3) Two-step aggregation: Unlike standard convolutions that apply a single transformation, ring aggregation performs two separate transformations—one within each ring and another across rings. In the second step, weighted averages can be applied to capture different spatial relationships (Figure 98).

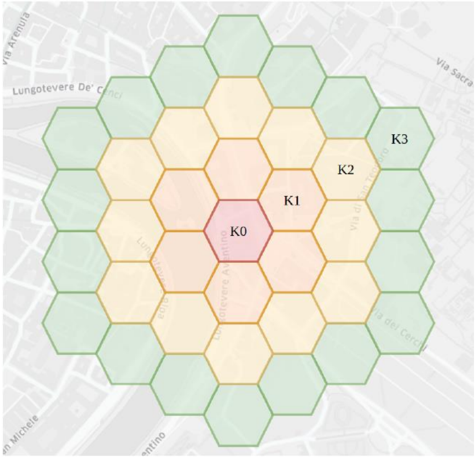


Figure 97: Spatial convolutions as developed by (Raczycki, 2021). A receptive field.

The two steps are as follows. First, for each spatial unit (hexagon), we aggregate (average) transformed concatenated embeddings from its neighbouring concentric rings (k-rings). Second, we apply another transformation to the aggregated result of each ring and then apply a weighted average across the rings to obtain the final embedding for the central spatial unit. The transformations before the aggregation process can be removed, removing the neural networks and making it a simple aggregation. Alternatively, when transformations are included, it is a learned aggregation.

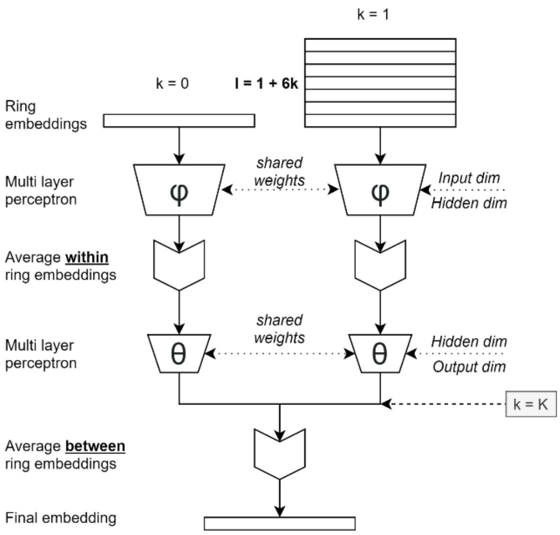


Figure 98: Learned ring aggregation fusion network.

7.3.2 Mathematical Formalism

The mathematical formulation of our ring aggregation strategy is as follows: denote the concatenated representation per hexagon as Ri, the mean of a ring as Mk, its weight as Wk, and the resultant central aggregated embedding as S (averaged over K rings), then the central embedding is expressed as a weighted sum across k-rings. Each k has a transformed average embedding per k-ring:

$$S = \sum_{k=0}^{K} W_k * f_\theta(M_k)$$

The average per ring equals the inverse of the number of hexagons in that ring and the sum over transformed concatenated representations of spatial units within that ring:

$$M_k = \frac{1}{I} * \sum_{i=1}^{I} f_\varphi(R_i)$$

Combined, this gives an expression with two learnable neural networks. Both networks map one representation to another, reducing the dimensionality at each step. The networks are parameterised by weights theta for across rings and phi for within. In total, this gives:

$$S = \sum_{k=0}^{K} W_k * f_\theta\left(\frac{\sum_{i=1}^{I} f_\varphi(R_i)}{I}\right)$$

7.3.3 Weighting Schemes

The contribution across rings is defined according to four different weighting schemes: natural exponent, logarithm, linear, and flat:

$$Exponential\ weighting\ W_k = e^{-k}$$

$$Logarithmic\ weighting\ W_k = \frac{1}{\log_2(k+2)}$$

$$Linear\ weighting\ W_k = 1 - \frac{1}{K}$$

$$Flat\ weighting\ W_k = \frac{1}{K}$$

While Raczycki (2021) employed the exponent, linear, and flat weightings, this study additionally includes a logarithmic weighted average. This addition aligns with the calculation of the physical living environment as operationalised by the Leefbaarometer.

7.4 Application

This chapter details the practical application of our novel urban representation learning methodology to assess liveability in the province of South Holland, Netherlands. We begin by describing our study area and the process of integrating Leefbaarometer scores with our spatial units. We then outline the diverse data sources employed and their preprocessing steps. The core of our application—the model architecture and training process—is then explained, including our innovative late-stage fusion approach using ring aggregation. We conclude by discussing our evaluation metrics and providing a comprehensive overview of how we operationalise and assess urban liveability through advanced machine-learning techniques.

7.4.1 Study Area and Leefbaarometer Scores

The province of South Holland in the Netherlands was selected as the study area for this research. Diverse urban landscapes, dense development, and a complex transportation network characterise this region. The selection of this study area allows for the prediction of Leefbaarometer scores across different spatial contexts. Spatial units in the province were filtered for the presence of Leefbaarometer scores and subsequently buffered to prepare padded rings for ring aggregation (Figure 99).

Figure 99: Buffered selection of spatial units.

Leefbaarometer scores were spatially joined with the hexagonal spatial units. The scores encompass five aspects: safety, social cohesion, physical environment, housing stock, and amenities, as well as the final liveability score. As the Leefbaarometer scores are provided in a grid of 100x100m cells, the spatial join with hexagonal units was based on overlapping surface area. Every H3 hexagon with at least a hectare of overlap with Leefbaarometer squares was assigned the six scores, weighted by the overlap of all matching squares. In Figure 7, spatial units with higher liveability scores are shown in red, while lower scores are depicted in blue (Figure 100).

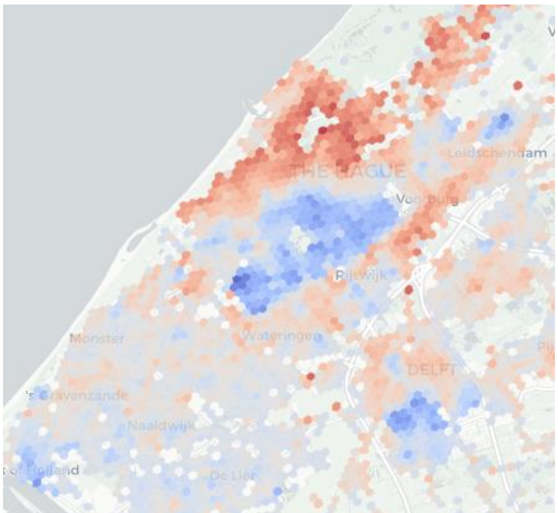


Figure 100: Spatially joined Leefbaarometer scores: a snippet of The Hague Westland region within the province of South Holland.

7.4.2 Data Sources and Preprocessing

The data sources employed in this study encompass a range of modalities, each providing a unique perspective on the urban environment. These include:

Road network data was obtained from OpenStreetMap, representing the car network and capturing proximity to highways and the potential exposure to the externalities of the transport system, like pollution of different areas within the region.

General Transit Feed Specification (GTFS) data retrieved from gtfs.org provides information about public transport schedules and routes, enabling the incorporation of public transport characteristics into urban representations.

Points of Interest (POI) data from OpenStreetMap offer insights into the distribution of amenities, parks and services across the region.

Street view images retrieved from Google Street View capture streetscapes' visual appearance and characteristics. Indexing of images follows the methodology developed by Garrido-Valenzuela et al. (2023)

Aerial images obtained from pdok.nl, an open dataset provided by the Dutch government, offer a bird's-eye view of the urban landscape and its physical characteristics. Each image has a 224 by 224 pixels resolution, which is required for image encoder neural networks.

7.4.3 Model Architecture and Training Process

Our model architecture consists of several components: feature extraction from data sources, late-fusion ring aggregation, and linear regression for valuation (Figure 101). This section will cover the loss function used to finetune the feature extractor for aerial images and the ring aggregation network. Subsequently, feature extraction and late-fusion using ring aggregation are outlined.

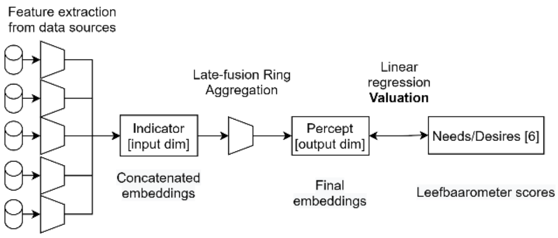


Figure 101: Modelling pipeline.

Loss Function

The loss function used to train the fusion network and finetune the image encoder network is a form of similarity loss. In particular, circle loss is applied since it is more likely to converge to optimal solutions than other forms of similarity loss (Y. Sun et al., 2020). Circle loss requires the sampling of triplets according to a heuristic. The fusion network and finetune have different heuristics, which will be detailed in the upcoming two sections.

Similarity loss takes duplicate neural networks, one for each sampled data point in the triplet (Figure 102). Representations of each data point are compared so that the difference between anchor and positive is minimised, whereas that between anchor and negative is maximised.

The calculated loss is then back-propagated through the duplicate neural network for the next batch of data.

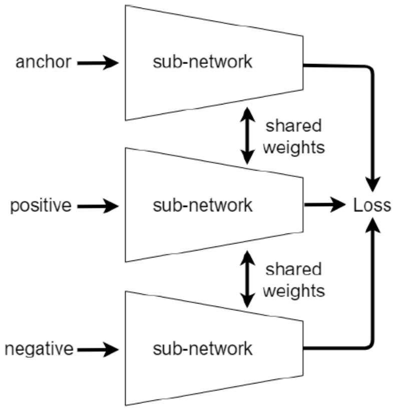


Figure 102: Triplet network. From (Ghojogh et al., 2022b).

Circle loss learns, as the name suggests, a circular decision boundary. It achieves this by weighing the deltas for positive and negative pairs based on the distance. In doing so, it corrects for exceeding imbalances in deltas, preventing minimising one at the expense of another.

$$Circle\ Loss(a, p, n) = log(1 + sum(exp(\gamma(\alpha\_p - \Delta\_p))) + sum(exp(\gamma(\Delta\_n - \alpha\_n))))$$

Feature Extraction

We use various encoders to create embeddings from our data sources (Table 4). All encoders, except for street view images, are trained or finetuned on data covering the buffered study area (Figure 99).

Table 4: Encoder networks used to create representations

Data	Encoder Network
Public transport stop characteristics	GTFS2vec (Gramacki et al., 2021)
Road network characteristics	Highway2vec (Leśniara & Szymański, 2022)
Aerial images	ConvNeXt Large (Liu et al., 2022)
Street-view images	ConvNeXt Large
Points of interest	Geovex (Donghi & Morvan, 2023) Hex2vec (Woźniak & Szymański, 2021)

Embeddings for all the data sources, except street view and aerial images, are obtained using a Python package called spatial representations for artificial intelligence (SRAI). Hence, the learning rates, batch sizes, and dimensionalities are all left to the default settings (Gramacki et al., 2023). The choice for these encoder networks is motivated by their availability in the Python package and suitability for the H3 hexagonal spatial index.

Street view images are fed into a large ConvNeXt model pre-trained using the ImageNet dataset (Russakovsky et al., 2015).

An image encoder is finetuned on aerial images. For this, we developed a novel sampling methodology based on location-based accessibility. Location-based accessibility applies the constrained maximum entropy principle (Hansen, 1972). It describes a dual process of obtaining the flattest possible probability distribution (entropy maximisation) while constraining it where applicable. In this context, constraints are applied closest to the centrally sampled hexagon. Further away from the centre, the flattening process gains the upper hand. Positive samples in the triplet relate to constraints, and negatives to maximum entropy—approximating negative mutual information.

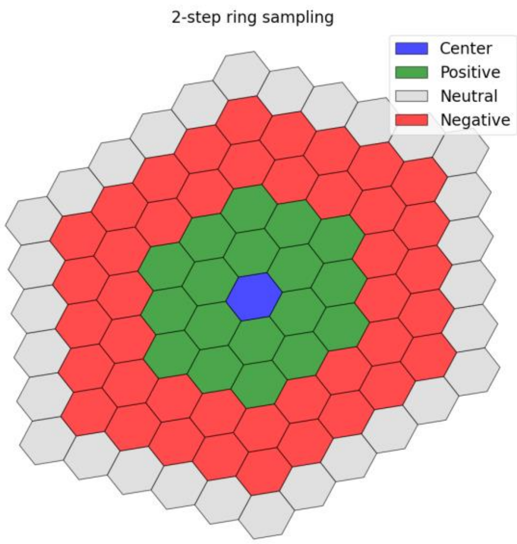


Figure 103: Two-step ring sampling used to finetune aerial image encoder network.

Concatenation

Next, the features of individual data sources are concatenated into a single embedding. During preliminary experimentation, we found that the relative dimensionality of embeddings from different data types matters greatly for final performance. Before concatenation, all extracted features that act as indicators should be sized roughly equally. Hence, we reduce the dimensionality of encoded images to 100 before concatenating using principal component analysis.

Late-Stage Fusion with Ring Aggregation

Our late-stage fusion approach utilises a learnable fusion network to integrate multi-modal urban embeddings, capturing spatial relationships through ring aggregation. This method builds upon the H3 hexagonal grid system, allowing for efficient representation of spatial context.

Central to our approach is considering inter-unit relationships, which form the basis for our sampling heuristic. We incorporate two primary measures of spatial relationship: Euclidean distance and location-based accessibility.

Euclidean distance is calculated between the centroids of spatial units. Given the relatively short distances within our study area, we do not account for the earth's curvature in these calculations. This simple metric provides a baseline measure of proximity between units.

Location-based accessibility, on the other hand, offers a more nuanced view of spatial relationships by combining travel times with a measure of destination attractiveness. We use OpenStreetMap walking, cycling and driving networks for travel times, linking spatial unit centroids to nodes in these respective networks for calculation using single source Dijkstra. For attractiveness, we use building density, specifically the Floor Space Index (FSI) from the RUDIFUN dataset (Harbers et al., 2022). The FSI captures the ratio of floor space to building footprint, including associated land uses like parking lots. More stories and less open space lead to higher FSI scores. We assign FSI scores to our hexagonal units using a weighted sum based on spatial overlap, mirroring our approach for Leefbaarometer scores.

The accessibility of each spatial unit is calculated:

$$A_i = \sum_j FSI_j * f(T_{ij})$$

Where A_i is the accessibility of unit i, FSI_j is the Floor Space Index of unit j, and $f(T_{ij})$ is an impedance function based on travel time. We employ an exponential decay function for $f(T_{ij})$:

$$f(T_{ij}) = e^{-\beta * Tij} = e^{-0.001 * Tij}$$

This function applies a decay rate to travel times, with a one-hour cutoff (3600 seconds). While we use a fixed β value of 0.001 in this study, future work could refine this by using empirically derived values for different population segments.

These inter-unit relationships inform our sampling heuristic for training the fusion network. We process triplets of neighbourhoods; each triplet contains 93 concatenated embeddings corresponding to the number of spatial units fitting into a receptive field of five rings. We select a positive sample from the top 2% closest neighbourhoods for any given anchor neighbourhood, using either Euclidean distance (lowest 2%) or location-based accessibility (highest 98%) as the similarity measure. The negative sample is randomly selected from the remaining neighbourhoods.

The Ring Aggregation Neural Network architecture is comprised of three modules: the WithinRingNN, the BetweenRingNN, and the overarching RingAggregationNN, see Figure 98. The input to the network is a set of concatenated embeddings representing different data modalities for each spatial unit within the receptive field. A batch normalisation layer is incorporated at the network's input to address the challenges of normalising data across different modalities, ensuring consistent scaling across diverse data types.

The WithinRingNN processes embeddings within each ring, consisting of a batch normalisation layer and two fully connected layers with a final hidden dimension 96. The BetweenRingNN processes the aggregated embeddings from each ring, which also consist of two fully connected layers, ultimately producing an output with a dimensionality of 30, matching the smallest input embedding data type dimension.

In both the WithinRingNN and BetweenRingNN, the common ReLU activation function is replaced with GeLU (Gaussian Error Linear Unit). This choice is motivated by GeLU's ability to prevent excessive sparsity in the learned representations, which proved crucial for improving the downstream performance in predicting Leefbaarometer scores. It seems that one should either go really dense or sparse.

The RingAggregationNN integrates these components in a forward pass that applies the WithinRingNN to all embeddings in each ring, pools the transformed embeddings within each ring, applies the BetweenRingNN to the pooled embeddings, calculates weights based on a chosen scheme (exponential, logarithmic, linear, or flat), and produces the final embedding through a weighted sum of the transformed ring embeddings.

The fusion network is trained using circle loss. The hyperparameters gamma and m are set to 250 and 0.15, respectively, as optimised by Sun et al. (2020). Utilising the Adam optimiser (Kingma & Ba, 2014), experiments demonstrated acceptable performance with a batch size of 256 and a learning rate of 0.0001.

7.4.4 Evaluation Metrics

We use a simple linear regression model to evaluate the performance of urban embeddings. Embeddings are used as the independent variable, and each liveability score is used as a dependent variable. Before doing so, the dimensionality of embeddings is reduced to 30 to ensure equal conditions. Preliminary experimentation showed that dimensionality strongly impacts the accuracy of the linear regression. The valuation measure is the R-squared score, which describes the explained variance.

Alternatively, Kendal tau could have been used since leefbaarometer scores are ordinal, as done by Levering et al. (2023). Instead, we assume that the relationship between Leefbaarometer scores and embeddings is linear, as dimensionality reduction is applied throughout this thesis using principal component analysis.

Experimentation of different data type combinations is done by training the ring aggregation model on all data sources. Then, the inference was performed on a zero-filled concatenated embedding. Therefore, the input of ring aggregation is mainly filled with zeros when comparing data sources.

Agglomerative clustering is used to compare results visually. Agglomerative clustering is hierarchical, aligning it with the hierarchical configuration of the H3 DGGS.

7.5 Results

This section presents the findings of our urban representation learning methodology, which is evaluated for its performance in predicting Leefbaarometer scores. Our results are structured around four key areas: sampling heuristics, the configuration of spatial convolutions (number of k-rings and weighting schemes), the influence of different data sources, and a comparison with existing studies in the field.

7.5.1 Sampling Heuristic

The choice of sampling heuristic determines which spatial units are selected as positives and negatives for a given anchor in our ring aggregation method. We compared two approaches: location-based accessibility and Euclidean distance.

Results indicate that the sampling heuristic has little impact on the predictive ability of the Leefbaarometer scores. Only the physical environment is affected, scoring better using Euclidean distance.

Table 5: Results, impact of sampling heuristic. Comparing location-based accessibility and Euclidean distance.

Score	Accessibility	Euclidean
Liveability	0.24	0.24
Amenities	0.71	0.71
Physical Environment	0.25	0.29
Social Cohesion	0.41	0.41
Safety	0.46	0.47

The cluster plot for location-based accessibility (Figure 12) reveals that larger urban areas, such as The Hague (light blue) and Rotterdam (red), are kept together in coherent clusters. Location-based accessibility captures functional relationships between areas, reflecting travel patterns and urban connectivity. In contrast, clusters for embeddings created using Euclidean distance (Figure 13) tend to break apart these larger urban areas.

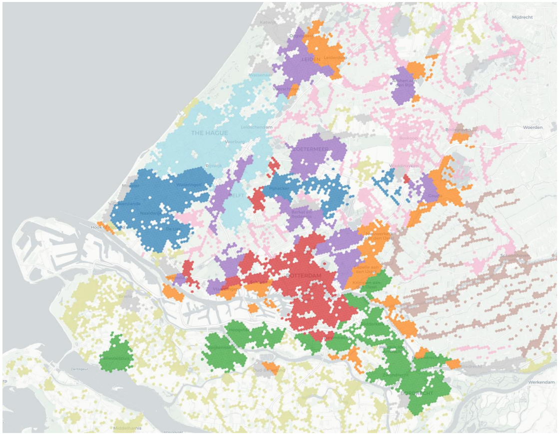


Figure 104: Ten agglomerative clusters for embeddings obtained using location-based accessibility.

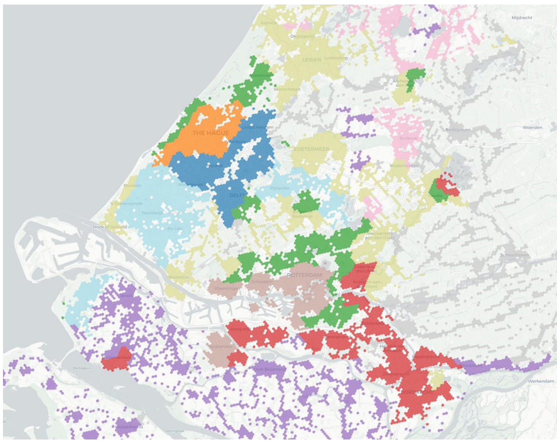


Figure 105: Ten agglomerative clusters for embeddings obtained using Euclidean distance.

7.5.2 Number of K-rings and Weighted Average

The configuration of the ring aggregation method, including the number of k-rings and weighting scheme, affects the quality of urban representations. Tables 4 and 5 present the best-performing configurations for simple and learnt aggregation.

For simple aggregation, linear weighting consistently performs best across all Leefbaarometer scores. The optimal number of k-rings varies; k=5 for liveability, physical environment, and safety; k=3 for amenities; and k=1 for social cohesion and housing stock. Different aspects of liveability operate at different spatial scales.

Table 6: Best-performing embeddings for simple aggregation.

Score	**K**	**Weight Type**	**R^2**
Liveability	5	linear	0.25
Amenities	3	linear	0.74
Physical Environment	5	linear	0.32
Social Cohesion	1	linear	0.46
Safety	5	linear	0.47
Housing Stock	1	linear	0.29

Learnt aggregation shows more variety in optimal configurations than simple aggregation. Notably, flatter weighting functions (logarithmic and flat) perform better for several scores, including liveability, amenities, physical environment, and safety. Social cohesion, however, benefits from the steeper linear weighting. The optimal number of k-rings for learnt aggregation is generally higher, with k=5 for most scores and k=3 for social cohesion.

Table 7: Best performing embeddings for learnt aggregation.

Score	K	Weight Type	R^2
Liveability	5	logarithm	0.26
Amenities	5	logarithm	0.72
Physical Environment	5	flat	0.29
Social Cohesion	3	linear	0.43
Safety	5	flat	0.46

Comparing the two methods, simple aggregation slightly outperforms learnt aggregation for most scores, except for liveability, where learnt aggregation shows a marginal improvement (R^2 of 0.26 vs 0.25). The performance difference is most noticeable for social cohesion (R^2 of 0.46 for simple vs 0.43 for learnt) and physical environment (R^2 of 0.32 for simple vs 0.29 for learnt).

The preference for flatter functions in learnt aggregation, contrasted with the consistent performance of the steeper linear function in simple aggregation, suggests that the learning process may identify relationships that benefit from a consideration of near and far contexts.

7.5.3 Data Sources

Data source analysis shows that images and points of interest have high predictive performance across Leefbaarometer scores. Finetuned aerial images outperform other modalities for liveability (0.34), social cohesion (0.48), housing stock (0.37), and safety (0.52) but underperform for physical environment and amenities. Street view and non-finetuned aerial images show balanced performance. Point-of-interest encoders perform well, with trade-offs between liveability and physical environment predictions. Road network and GTFS embeddings excel in predicting amenities, safety, and social cohesion but score lower on physical environment and liveability. Simple aggregation outperforms learned aggregation for all data sources, with the difference most pronounced for aerial image embeddings.

7.5.4 Comparison to Other Studies

Comparison with other studies shows incremental improvements across three steps: street-view images, addition of POI data, and incorporation of proximity measures. Urban2Vec Wang et al. (2020) implement the first two steps, while Huang et al. (2021) add the third. Predictive accuracy improves with each step, with the largest gains in amenities, social cohesion, and safety scores. Euclidean distance and accessibility as proximity measures perform similarly. The ring aggregation method developed in this study outperforms the compared approaches across all Leefbaarometer scores, with notable improvements in amenities (0.85 vs 0.64), social cohesion (0.6 vs 0.43), and overall liveability (0.35 vs 0.27).

Table 8: Impact of data sources on the predictive accuracy of Leefbaarometer scores. Simple ring aggregation / learnt ring aggregation using Euclidean distance.

Score	GeoVex	Hex2Vec	Aerial (finetuned)	Aerial (pre-trained)	Road Network	GTFS	Street View
Liveability	0.24/0.24	0.17/0.16	0.34/0.11	0.23/0.098	0.12/0.12	0.16/0.11	0.22/0.1
Amenities	0.73/0.73	0.74/0.7	0.68/0.47	0.64/0.5	0.63/0.63	0.69/0.7	0.72/0.58
Physical Environment	0.29/0.27	0.35/0.32	0.27/0.11	0.3/0.12	0.2/0.2	0.08/0.08	0.26/0.067
Social Cohesion	0.42/0.42	0.45/0.44	0.48/0.24	0.45/0.28	0.33/0.33	0.36/0.38	0.42/0.29
Safety	0.46/0.45	0.49/0.48	0.52/0.35	0.45/0.31	0.34/0.34	0.42/0.44	0.47/0.35
Housing Stock	0.26/0.27	0.25/0.24	0.37/0.18	0.29/0.17	0.19/0.19	0.21/0.23	0.25/0.19

Table 9: Comparison to other studies.

Leefbaarometer Score	Step 1	Step 2	Step 3 (Euclidean)	Step 3 (Accessibility)	Ring Aggregation
Liveability	0.2	0.21	0.24	0.27	0.35
Amenities	0.41	0.46	0.64	0.62	0.85
Physical Environment	0.19	0.21	0.28	0.29	0.4
Social Cohesion	0.35	0.41	0.43	0.43	0.6
Safety	0.3	0.34	0.43	0.43	0.54

7.6 Conclusion

This study has successfully demonstrated the potential of urban representation learning to automate the operationalisation of liveability, addressing the challenges of resource-intensive and time-consuming traditional methods. Our novel ring aggregation method leverages spatial convolutions on the H3 hexagonal DGGS. By utilising deep neural networks and diverse data sources—including aerial and street view images, road network characteristics, public transport data, and points of interest—we have created a method capable of capturing the multifaceted nature of urban environments and their liveability.

Our sensitivity analysis revealed insights into the optimal configuration of the ring aggregation method, demonstrating that different aspects of liveability operate most effectively at varying sizes of receptive fields. This finding underscores the complex spatial dynamics of liveability and the importance of multi-scale approaches in urban analysis. Furthermore, our results highlight the significant impact of data source selection on predictive performance, with finetuned aerial images outperforming other modalities for several liveability aspects. These insights emphasise the importance of carefully considering and combining diverse data sources in urban representation learning to capture the full spectrum of liveability scores.

The ring aggregation method developed in this study has shown substantial improvements over existing urban representation learning approaches across all Leefbaarometer scores, with notable advancements in predicting amenities, social cohesion, and overall liveability. These results validate our approach's effectiveness and demonstrate its potential to enhance our understanding and assessment of urban liveability. By automating the creation and perception of indicators, this research contributes to more efficient and data-driven policy evaluation in urban planning, providing valuable tools for policymakers and planners to quickly and accurately assess liveability across different urban contexts.

7.7 Recommendations

While our study has made strides in advancing the field of urban representation learning and its application to liveability assessment, it has also revealed areas for further exploration and improvement. Building on our findings and acknowledging the current limitations of our approach, we propose two main directions for future work.

First, the model architecture could be improved. While effective, our current ring aggregation method may be limited in capturing highly localised or ring-specific features due to using a single shared neural network for all within and across ring transformations. Learned ring aggregation performed worse than simple aggregation, which may be explained by the bottleneck effect (Alon & Yahav, 2021). More complex architectures may be necessary to mitigate this effect and capture fine-grained spatial information, such as convolutional networks designed explicitly for hexagonal grids (Donghi & Morvan, 2023; Hoogeboom et al., 2018). Future research could explore fusion networks with diverse data sources, incorporating additional modalities such as text data, sound measurements, demographics, building characteristics, and 3D or pollution (volumetric) data to create richer urban representations.

Miller et al. (2013b) have developed an operationalisation of liveability focussed on transportation, noting that spatial resolution—granularity—significantly impacts performance. Granularity aligns with the number of spatial units (H3 hexagons), as described in their work on granular computing by Wilke & Portmann (2016).

The H3 DGGS is hierarchically structured. Future work can take advantage of this structure by considering the following two points. 1) Designing urban representation learning methodology incorporating travel trajectories to learn hierarchical embeddings (Chen et al., 2023; Hu et al., 2023). 2) Hierarchical dynamic generative modelling, operationalised using renormalising generative models (K. Friston et al., 2024).

Furthermore, granular computing may capture the reciprocal (hierarchical) relationship which defines liveability. See Figure 106 for all possible states bridging the relationship between two granules. Figure 107 shows a sampling of this total set, either two or three times. Adding a third (d) gives the model hierarchical layers or temporal depth—like looking towards the vanishing point in a room with dark walls and low white ceilings, Q floats in the middle of that room, coming up from the paper towards the viewer. Temporal depth allows for the operationalisation of sustainability in addition to liveability. Hence, a temporally deep environment-resident relationship can be operationalised in granular geometry. See Figure 93: Reciprocal relationship between human/resident and environment. Perspectival measurement.

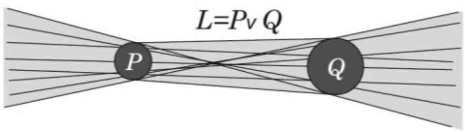


Figure 106: A construction heuristic for the granular line connection of two granular points. From (Wilke & Portmann, 2016).

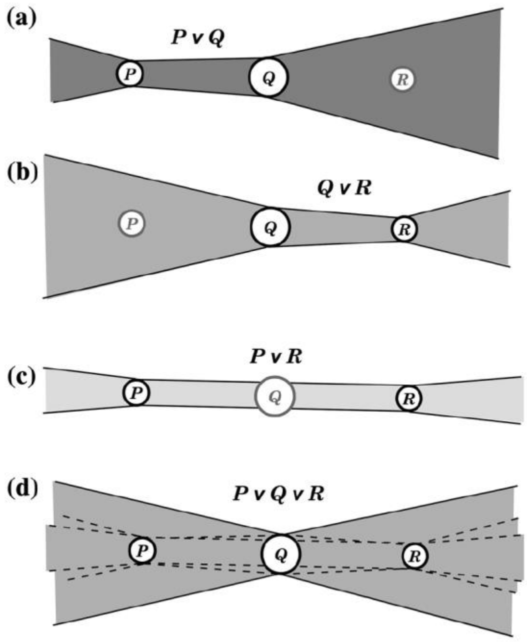


Figure 107: Heuristic line connection a-c of two points, d of three points. From (Wilke & Portmann, 2016).

The second direction for future work involves reconceptualising the assessment of liveability itself. While our current approach has shown promise in automating the operationalisation of liveability, it is based on the Leefbaarometer, which has inherent limitations as a measurement model rather than a causal one (Mandemakers et al., 2021). The main one is that it should not be used to forecast the effect of interventions in the urban environment on liveability. To move beyond these limitations, we suggest framing the resident-environment relationship as transactional (Aitken & Bjorklund, 1988), viewing liveability as a dynamic process rather than a static outcome. This perspective aligns with previous recommendations from Leidelmeijer (2004) for a dynamic approach to liveability studies and could be operationalised using generative models (M. J. D. Ramstead et al., 2019). Such an approach would allow for the incorporation of individual variations in perception and needs/desires, moving away from the assumption of uniformity implicit in current liveability measures and potentially offering a more comprehensive and nuanced understanding of urban liveability.

7.8 References

Aitken, S. C., & Bjorklund, E. M. (1988). TRANSACTIONAL AND TRANSFORMATIONAL THEORIES IN BEHAVIORAL GEOGRAPHY. The Professional Geographer, 40(1), 54–64. https://doi.org/10.1111/j.0033-0124.1988.00054.x

Chen, N., Yang, A., Chen, L., Xiong, W., & Jing, N. (2023). STO2Vec: A Multiscale Spatio-Temporal Object Representation Method for Association Analysis. ISPRS International Journal of Geo-Information, 12(5), 207. https://doi.org/10.3390/ijgi12050207

Donghi, D., & Morvan, A. (2023). GeoVeX: Geospatial Vectors with Hexagonal Convolutional Autoencoders. Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 3–13. https://doi.org/10.1145/3615886.3627750

Dorst, M. van. (2005). Een duurzaam leefbare woonomgeving: Fysieke voorwaarden voor privacyregulering. Eburon.

Fan, Z., Zhang, F., Loo, B. P. Y., & Ratti, C. (2023). Urban visual intelligence: Uncovering hidden city profiles with street view images. Proceedings of the National Academy of Sciences, 120(27), e2220417120. https://doi.org/10.1073/pnas.2220417120

Friston, K., Heins, C., Verbelen, T., Costa, L. D., Salvatori, T., Tschantz, A., Koudahl, M., Buckley, C., & Parr, T. (2024). From pixels to planning: Scale-free active inference.

Garrido-Valenzuela, F., Cats, O., & Van Cranenburgh, S. (2023). Where are the people? Counting people in millions of street-level images to explore associations between people's urban density and urban characteristics. Computers, Environment and Urban Systems, 102, 101971. https://doi.org/10.1016/j.compenvurbsys.2023.101971

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2022). Spectral, Probabilistic, and Deep Metric Learning: Tutorial and Survey (arXiv:2201.09267). arXiv. http://arxiv.org/abs/2201.09267

Gramacki, P., Leśniara, K., Raczycki, K., Woźniak, S., Przymus, M., & Szymański, P. (2023). SRAI: Towards Standardization of Geospatial AI. Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 43–52. https://doi.org/10.1145/3615886.3627740

Gramacki, P., Woźniak, S., & Szymański, P. (2021). gtfs2vec—Learning GTFS Embeddings for comparing Public Transport Offer in Microregions. https://doi.org/10.48550/ARXIV.2111.00960

Grover, A., & Leskovec, J. (2016). node2vec: Scalable Feature Learning for Networks (arXiv:1607.00653). arXiv. http://arxiv.org/abs/1607.00653

Gupta, J., Chen, Y., & Rammelt, C. (2024). Transport within earth system boundaries. Npj Sustainable Mobility and Transport, 1(1), 5. https://doi.org/10.1038/s44333-024-00005-5

Hansen, S. (1972). Utility, Accessibility and Entropy in Spatial Modelling. The Swedish Journal of Economics, 74(1), 35. https://doi.org/10.2307/3439008

Harbers, A., van Amsterdam, H., & Spoon, M. (2022). RUDIFUN 2022: Ruimtelijke Dichtheden en Functiemenging in Nederland.

Hoogeboom, E., Peters, J. W. T., Cohen, T. S., & Welling, M. (2018). HexaConv (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1803.02108

Hu, S., Gao, S., Luo, W., Wu, L., Li, T., Xu, Y., & Zhang, Z. (2023). Revealing intra-urban hierarchical spatial structure through representation learning by combining road network abstraction model and taxi trajectory data. Annals of GIS, 29(4), 499–516. https://doi.org/10.1080/19475683.2023.2241526

Huang, T., Wang, Z., Sheng, H., Ng, A. Y., & Rajagopal, R. (2021). Learning Neighborhood Representation from Multi-Modal Multi-Graph: Image, Text, Mobility Graph and Beyond (arXiv:2105.02489). arXiv. http://arxiv.org/abs/2105.02489

Kim, N., & Yoon, Y. (2022). Effective Urban Region Representation Learning Using Heterogeneous Urban Graph Attention Network (HUGAT) (arXiv:2202.09021). arXiv. http://arxiv.org/abs/2202.09021

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization (Version 9). arXiv. https://doi.org/10.48550/ARXIV.1412.6980

Kipf, T. N., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks (arXiv:1609.02907). arXiv. http://arxiv.org/abs/1609.02907

Leidelmeijer, K. (2004). RIVM rapport 630950002 Kwaliteit van de Leefomgeving en Leefbaarheid.

Leśniara, K., & Szymański, P. (2022). highway2vec: Representing OpenStreetMap microregions with respect to their road network characteristics. Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 18–29. https://doi.org/10.1145/3557918.3565865

Levering, A., Marcos, D., Van Vliet, J., & Tuia, D. (2023). Predicting the liveability of Dutch cities with aerial images and semantic intermediate concepts. Remote Sensing of Environment, 287, 113454. https://doi.org/10.1016/j.rse.2023.113454

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2201.03545

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

Miller, H. J., Witlox, F., & Tribby, C. P. (2013). Developing context-sensitive livability indicators for transportation planning: A measurement framework. Journal of Transport Geography, 26, 51–64. https://doi.org/10.1016/j.jtrangeo.2012.08.007

Moschella, L., Maiorca, V., Fumero, M., Norelli, A., Locatello, F., & Rodolà, E. (2023). Relative representations enable zero-shot latent space communication (arXiv:2209.15430). arXiv. http://arxiv.org/abs/2209.15430

Pacione, M. (2003). Urban environmental quality and human wellbeing—A social geographical perspective. Landscape and Urban Planning, 65(1–2), 19–30. https://doi.org/10.1016/S0169-2046(02)00234-7

Raczycki, K. (2021). Predicting the Location of Bicycle-sharing Stations using OpenStreetMap Data (arXiv:2111.01722). arXiv. http://arxiv.org/abs/2111.01722

Ramstead, M. J. D., Constant, A., Badcock, P. B., & Friston, K. J. (2019). Variational ecology and the physics of sentient systems. Physics of Life Reviews, 31, 188–205. https://doi.org/10.1016/j.plrev.2018.12.002

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge (arXiv:1409.0575). arXiv. http://arxiv.org/abs/1409.0575

Sun, F., Qi, J., Chang, Y., Fan, X., Karunasekera, S., & Tanin, E. (2023). Urban Region Representation Learning with Attentive Fusion (arXiv:2312.04606). arXiv. http://arxiv.org/abs/2312.04606

Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z., & Wei, Y. (2020). Circle Loss: A Unified Perspective of Pair Similarity Optimization (arXiv:2002.10857). arXiv. http://arxiv.org/abs/2002.10857

van Cranenburgh, S., & Garrido-Valenzuela, F. (2023). Computer vision-enriched discrete choice models, with an application to residential location choice.

Veenhoven, R. (2000). LEEFBAARHEID Betekenissen en meetmethoden.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph Attention Networks (arXiv:1710.10903). arXiv. http://arxiv.org/abs/1710.10903

Wang, Z., Li, H., & Rajagopal, R. (2020). Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding. Proceedings of the AAAI Conference on Artificial Intelligence, 34(01), 1013–1020. https://doi.org/10.1609/aaai.v34i01.5450

Wilke, G., & Portmann, E. (2016). Granular computing as a basis of human–data interaction: A cognitive cities use case. Granular Computing, 1(3), 181–197. https://doi.org/10.1007/s41066-016-0015-4

Woźniak, S., & Szymański, P. (2021). Hex2vec—Context-Aware Embedding H3 Hexagons with OpenStreetMap Tags. Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, 61–71. https://doi.org/10.1145/3486635.3491076

Xiao, L., Lo, S., Zhou, J., Liu, J., & Yang, L. (2021). Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. Environment and Planning B: Urban Analytics and City Science, 48(8), 2363–2384. https://doi.org/10.1177/2399808320977866

Zhang, M., Li, T., Li, Y., & Hui, P. (2020). Multi-View Joint Graph Representation Learning for Urban Region Embedding. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 4431–4437. https://doi.org/10.24963/ijcai.2020/611

Appendix B: Technical Details

The implementation of our two learning strategies revealed several nuances that became apparent only during the execution phase. These include unexpected training durations for certain models and potential overtraining concerns, particularly in the case of learning strategy two.

Activation Function Learning Strategy One

After preliminary testing, it was decided to exchange ReLU for GeLU since the latter has a smoother transition near 0, meaning there is less sparsity in the embeddings. For dense representations, sparsity is not beneficial. Since there are few layers where activation is applied, this leads to pronounced dead spots in the embeddings.

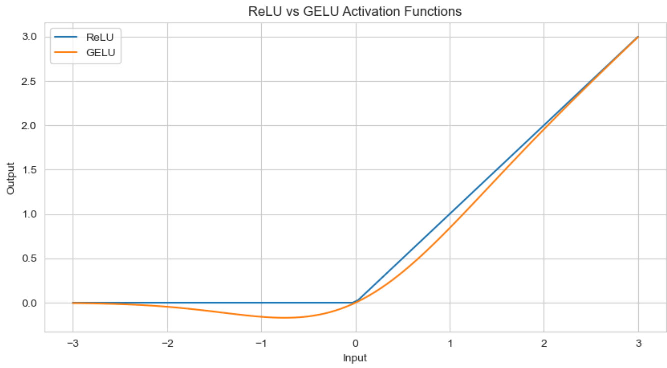


Figure 108: Comparison of activation functions. ReLU versus GELU. Rectified Linear Unit and Gaussian Error Linear Unit.

Computational Details Learning Strategy Two

Step 1: Finetuning ConvNeXt Model

The first step involved finetuning a large ConvNeXt model with approximately 200 million parameters. At H3 resolution 10, convergence of circle loss at gamma ~250 was achieved in two epochs using 240,000 images. At resolution 9 the training required twelve epochs. This difference in convergence time makes sense, given the ratio of hexagons increasing by a factor of 6 at each k-ring. The formula for the number of hexagons per ring is 1+k*6, explaining the similar total compute required for a given loss across resolutions.

Step 2: Skip-gram-like Models

The second step involved training two skip-gram-like models: one for spatial unit embeddings and another for POI indices. This process is akin to creating two lookup tables of embeddings using torch.nn.Embedding(), took approximately four hours to run. Despite the models containing around 13 million parameters (with embeddings of dimension 1536 for each spatial unit and POI index), the computation was relatively quick due to the small size of the input data set. The main bottleneck was memory speed rather than computational power, leading us to use CPU instead of GPU to minimise data transfers over PCI.

Training Curves and Considerations

Our training runs consistently followed tightly distributed logistic curves. Early stopping is implemented to avoid catastrophic forgetting of results from previous steps. If continued, embeddings collapse into fully fitting the current step and become unusable.

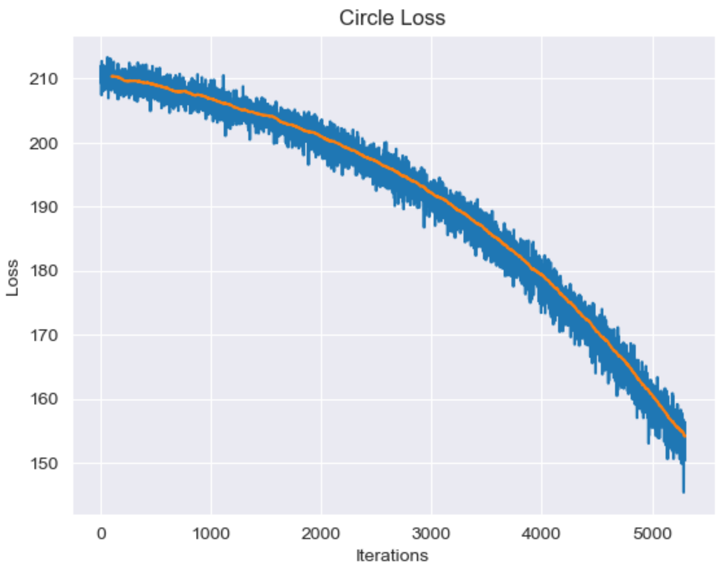


Figure 109: Every training run of steps 2 and 3 follows a tightly distributed logistic curve. It is necessary to break off training halfway through this curve to prevent training from previous steps. Training loss shown is that of step 3 as learned for H3 resolution 9.

Finetuning encoder networks is considerably more expensive than training a smaller fusion network on concatenated views of several smaller pre-trained encoders. For example, tuning ConvNeXt large on an RTX 3090 took about ten hours compared to just minutes for the fusion network.

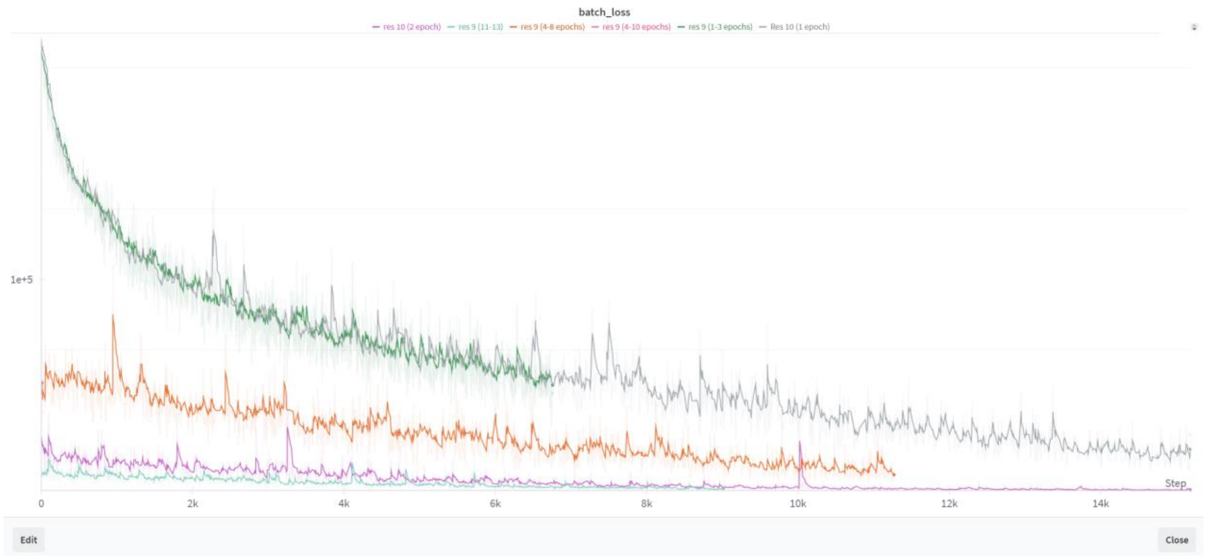


Figure 110: Circle loss (log scale) of finetuning two ConvNeXt models. One for H3 resolution 9 and another for H3 resolution 10. Resolution 9 has about six times less hexagons than 10. Hence, the difference in the number of epochs is thirteen versus two. Ten thousand steps of 16 hexagons/images per batch is about 5 hours of finetuning time on an Rtx 3090.

Appendix C: Point of Interest Dictionary

Both learning strategies one and two have some form of point of interest data. The GEOFABRIK filter provided in the SRAI Python package was used and then filtered for values present in the study area. The following list is an overview of keys used in step two of learning strategy two. In practice, there are some differences between H3 resolutions, but they are not very important for demonstration purposes (about ten POI keys difference).

aerialway=magic_carpet
aerialway=platter
aerialway=rope_tow
aerialway=zip_line
aeroway=aerodrome
aeroway=apron
aeroway=helipad
amenity=arts_centre
amenity=atm
amenity=bank
amenity=bar
amenity=bench
amenity=bicycle_parking
amenity=bicycle_rental
amenity=biergarten
amenity=bus_station
amenity=cafe
amenity=car_rental
amenity=car_sharing
amenity=car_wash
amenity=cinema
amenity=clinic
amenity=college
amenity=community_centre
amenity=courthouse
amenity=dentist
amenity=doctors
amenity=drinking_water
amenity=fast_food
amenity=ferry_terminal
amenity=fire_station
amenity=food_court
amenity=fountain
amenity=fuel
amenity=grave_yard
amenity=hospital
amenity=kindergarten
amenity=library
amenity=marketplace
building=office
building=outbuilding
building=parking
building=pavilion
building=public
building=residential
building=retail
building=roof
building=ruins
building=school
building=semidetached_house
building=service
building=shed
building=silo
building=sports_centre
building=stable
building=static_caravan
building=storage_tank
building=supermarket
building=temple
building=terrace
building=toilets
building=train_station
building=transformer_tower
building=transportation
building=university
building=warehouse
building=yes
emergency=fire_hydrant
emergency=phone
foot=designated
highway=bridleway
highway=bus_stop
highway=busway
highway=crossing
highway=cycleway
highway=emergency_access_point
highway=footway
highway=living_street
leisure=slipway
leisure=sports_centre
leisure=stadium
leisure=swimming_pool
leisure=water_park
man_made=lighthouse
man_made=pier
man_made=surveillance
man_made=tower
man_made=wastewater_plant
man_made=water_tower
man_made=water_well
man_made=water_works
man_made=watermill
man_made=windmill
natural=beach
natural=heath
natural=peak
natural=spring
natural=tree
natural=water
natural=wetland
natural=wood
office=diplomatic
public_transport=stop_position
railway=halt
railway=level_crossing
railway=light_rail
railway=miniature
railway=monorail
railway=narrow_gauge
railway=rail
railway=station
railway=subway
railway=tram
railway=tram_stop
shop=alcohol
shop=bakery
shop=beauty

amenity=nightclub
amenity=nursing_home
amenity=parking
amenity=pharmacy
amenity=place_of_worship
amenity=police
amenity=post_box
amenity=post_office
amenity=prison
amenity=pub
amenity=public_building
amenity=recycling
amenity=restaurant
amenity=school
amenity=shelter
amenity=taxi
amenity=telephone
amenity=theatre
amenity=toilets
amenity=townhall
amenity=university
amenity=vending_machine
amenity=veterinary
amenity=waste_basket
boundary=national_park
building=allotment_house
building=apartments
building=barn
building=boathouse
building=bungalow
building=bunker
building=cabin
building=carport
building=chapel
building=church
building=civic
building=college
building=commercial
building=construction
building=cowshed
building=detached
building=dormitory
building=farm
building=farm_auxiliary
building=fire_station
building=garage
highway=mini_roundabout
highway=motorway
highway=motorway_junction
highway=motorway_link
highway=path
highway=pedestrian
highway=primary
highway=primary_link
highway=residential
highway=road
highway=secondary
highway=secondary_link
highway=service
highway=services
highway=speed_camera
highway=steps
highway=stop
highway=street_lamp
highway=tertiary
highway=tertiary_link
highway=track
highway=traffic_signals
highway=trunk
highway=trunk_link
highway=turning_circle
highway=unclassified
historic=archaeological_site
historic=castle
historic=fort
historic=memorial
historic=monument
historic=ruins
historic=wayside_shrine
horse=designated
landuse=allotments
landuse=cemetery
landuse=commercial
landuse=farmland
landuse=farmyard
landuse=forest
landuse=grass
landuse=industrial
landuse=meadow
landuse=military
landuse=orchard
landuse=quarry
shop=beverages
shop=bicycle
shop=books
shop=butcher
shop=car
shop=car_repair
shop=chemist
shop=clothes
shop=computer
shop=convenience
shop=department_store
shop=doityourself
shop=dry_cleaning
shop=florist
shop=furniture
shop=garden_centre
shop=general
shop=gift
shop=greengrocer
shop=hairdresser
shop=hardware
shop=jewelry
shop=kiosk
shop=laundry
shop=mall
shop=mobile_phone
shop=newsagent
shop=optician
shop=outdoor
shop=shoes
shop=sports
shop=stationery
shop=supermarket
shop=toys
shop=travel_agency
shop=video
sport=swimming
sport=tennis
tourism=artwork
tourism=attraction
tourism=camp_site
tourism=caravan_site
tourism=chalet
tourism=guest_house
tourism=hostel
tourism=hotel

building=garages
building=ger
building=government
building=grandstand
building=greenhouse
building=hangar
building=hospital
building=hotel
building=house
building=houseboat
building=hut
building=industrial
building=kindergarten
building=kiosk
building=mosque

landuse=recreation_ground
landuse=reservoir
landuse=residential
landuse=retail
landuse=scrub
landuse=vineyard
leisure=common
leisure=dog_park
leisure=golf_course
leisure=ice_rink
leisure=marina
leisure=nature_reserve
leisure=park
leisure=playground
leisure=recreation_ground

tourism=information
tourism=motel
tourism=museum
tourism=picnic_site
tourism=theme_park
tourism=viewpoint
tourism=zoo
waterway=canal
waterway=dam
waterway=dock
waterway=drain
waterway=lock_gate
waterway=river
waterway=stream
waterway=waterfall
waterway=weir

References

Acker, V. V., & Witlox, F. (2008). *THEORIZING ON TRAVEL BEHAVIOR USING SOCIAL PSYCHOLOGY AND LIFESTYLE THEORY*.

Aitken, S. C., & Bjorklund, E. M. (1988). TRANSACTIONAL AND TRANSFORMATIONAL THEORIES IN BEHAVIORAL GEOGRAPHY. *The Professional Geographer*, *40*(1), 54–64. https://doi.org/10.1111/j.0033-0124.1988.00054.x

Albarracin, M., Hipólito, I., Tremblay, S. E., Fox, J. G., René, G., Friston, K., & Ramstead, M. J. D. (2023). *Designing explainable artificial intelligence with active inference: A framework for transparent introspection and decision-making*.

Albarracin, M., Pitliya, R. J., Ramstead, M. J. D., & Yoshimi, J. (2023). Mapping Husserlian Phenomenology onto Active Inference. In C. L. Buckley, D. Cialfi, P. Lanillos, M. Ramstead, N. Sajid, H. Shimazaki, & T. Verbelen (Eds.), *Active Inference* (Vol. 1721, pp. 99–111). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-28719-0_7

Albarracin, M., Ramstead, M., Pitliya, R. J., Hipolito, I., Da Costa, L., Raffa, M., Constant, A., & Manski, S. G. (2024). Sustainability under Active Inference. *Systems*, *12*(5), 163. https://doi.org/10.3390/systems12050163

Alex Pentland, & Pentland, A. (2020). Diversity of Idea Flows and Economic Growth. *Journal of Social Computing*, *1*(1), 71–81. https://doi.org/10.23919/jsc.2020.0007

Almaatouq, A., Prieto-Castrillo, F., & Pentland, A. (2016). Mobile Communication Signatures of Unemployment. In E. Spiro & Y.-Y. Ahn (Eds.), *Social Informatics* (Vol. 10046, pp. 407–418). Springer International Publishing. https://doi.org/10.1007/978-3-319-47880-7_25

Anciaes, P., & Jones, P. (2020). Transport policy for liveability – Valuing the impacts on movement, place, and society. *Transportation Research Part A: Policy and Practice*, *132*, 157–173. https://doi.org/10.1016/j.tra.2019.11.009

Annema, J. A., & Koopmans, C. (2015). The practice of valuing the environment in cost-benefit analyses in transport and spatial projects. *Journal of Environmental Planning and Management*, *58*(9), 1635–1648. https://doi.org/10.1080/09640568.2014.941975

Annema, J. A., Mouter, N., & Razaei, J. (2015). Cost-benefit Analysis (CBA), or Multi-criteria Decision-making (MCDM) or Both: Politicians' Perspective in Transport Policy Appraisal. *Transportation Research Procedia*, *10*, 788–797. https://doi.org/10.1016/j.trpro.2015.09.032

Appleyard, D. (1987). *Toward an Urban Design Manifesto*.

Avel Guénin—Carlut. (2022). Thinking like a State: Embodied intelligence in the deep history of our collective mind. *IOP Conference Series*, *1261*(1), 012026–012026. https://doi.org/10.1088/1757-899x/1261/1/012026

Balestriero, R., & LeCun, Y. (2024). *Learning by Reconstruction Produces Uninformative Features For Perception* (arXiv:2402.11337). arXiv. http://arxiv.org/abs/2402.11337

Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, *15*(2), 73–80. https://doi.org/10.1016/j.tranpol.2007.10.005

Banister, D., & Hickman, R. (2013). Transport futures: Thinking the unthinkable. *Transport Policy*, *29*, 283–293. https://doi.org/10.1016/j.tranpol.2012.07.005

Barker, R. G. (with Kansas). (1978). *Habitats, environments, and human behavior* (1st ed). Jossey-Bass.

Bejan, A., & Lorente, S. (2010). The constructal law of design and evolution in nature. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1545), 1335–1347. https://doi.org/10.1098/rstb.2009.0302

Beni, M. D., & Friston, K. (2024). Carving teleology at its joints. *Synthese*, *204*(1), 11. https://doi.org/10.1007/s11229-024-04660-6

Bouizegarene, N., Ramstead, M. J. D., Constant, A., Friston, K. J., & Kirmayer, L. J. (2024). Narrative as active inference: An integrative account of cognitive and social functions in adaptation. *Frontiers in Psychology*.

Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., & Colton, S. (2012). A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, *4*(1), 1–43. https://doi.org/10.1109/TCIAIG.2012.2186810

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417–2444. https://doi.org/10.1007/s11229-016-1239-1

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00599

Bruineberg, J., Rietveld, E., Parr, T., Thomas Parr, van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology*, *455*, 161–178. https://doi.org/10.1016/j.jtbi.2018.07.002

Chan, W., & Ren, Q. (2023). *Region-Wise Attentive Multi-View Representation Learning for Urban Region Embeddings* (arXiv:2307.03212). arXiv. http://arxiv.org/abs/2307.03212

Chen, N., Yang, A., Chen, L., Xiong, W., & Jing, N. (2023). STO2Vec: A Multiscale Spatio-Temporal Object Representation Method for Association Analysis. *ISPRS International Journal of Geo-Information*, *12*(5), 207. https://doi.org/10.3390/ijgi12050207

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Coffeng, G. (2018). *Schaalsprong openbaar vervoer Den Haag en regio*.

Constant, A., Clark, A., & Friston, K. J. (2021). Representation Wars: Enacting an Armistice Through Active Inference. *Frontiers in Psychology*, *11*, 598733. https://doi.org/10.3389/fpsyg.2020.598733

Constant, A., Clark, A., Kirchhoff, M., & Friston, K. J. (2022). Extended active inference: Constructing predictive cognition beyond skulls. *Mind & Language*, *37*(3), 373–394. https://doi.org/10.1111/mila.12330

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of The Royal Society Interface*, *15*(141), 20170685. https://doi.org/10.1098/rsif.2017.0685

Cordes, R. (2024). *Comments Submitted by BlockScience, University of Washington APL Information Risk and Synthetic Intelligence Research Initiative (IRSIRI), Cognitive Security and Education Forum (COGSEC), and the Active Inference Institute (AII) to the Networking and Information Technology Research and Development National Coordination Office's Request for Comment on The Creation of a National Digital Twins R&D Strategic Plan NITRD-2024-13379*. Zenodo. https://doi.org/10.5281/ZENODO.13273682

de Abreu e Silva, J., Currans, K. M., Van Acker, V., & Schneider, R. J. (Eds.). (2023). *Handbook on transport and land use: A holistic approach in an age of rapid technological change*. Edward Elgar Publishing.

de Boer, L., Mouter, N., & de vries, M. (2023). *Brede welvaart heeft ethische blik nodig*.

de Gruijter, M. (2019). *Mobility as a Service in the Paleiskwartier: Explanatory factors for the intention to use and a process evaluation*.

De Haan, F. J., Ferguson, B. C., Adamowicz, R. C., Johnstone, P., Brown, R. R., & Wong, T. H. F. (2014). The needs of society: A new understanding of transitions, sustainability and liveability. *Technological Forecasting and Social Change*, *85*, 121–132. https://doi.org/10.1016/j.techfore.2013.09.005

De Vos, J., Derudder, B., Van Acker, V., & Witlox, F. (2012). Reducing car use: Changing attitudes or relocating? The influence of residential dissonance on travel behavior. *Journal of Transport Geography*, *22*, 1–9. https://doi.org/10.1016/j.jtrangeo.2011.11.005

Dekker, T., Koster, P., & Mouter, N. (2019). *The economics of participatory value evaluation*. Tinbergen Institute.

Dettmers, T., Lewis, M., Shleifer, S., & Zettlemoyer, L. (2021). *8-bit Optimizers via Block-wise Quantization* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2110.02861

Dietrich, F., & List, C. (2016). *Mentalism Versus Behaviourism in Economics: A Philosophy-of-Science Perspective*. 36.

Donghi, D., & Morvan, A. (2023). GeoVeX: Geospatial Vectors with Hexagonal Convolutional Autoencoders. *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 3–13. https://doi.org/10.1145/3615886.3627750

Dorst, M. van. (2005). *Een duurzaam leefbare woonomgeving: Fysieke voorwaarden voor privacyregulering*. Eburon.

Downey, G. (2016). Being Human in Cities: Phenotypic Bias from Urban Niche Construction. *Current Anthropology*, *57*(S13), S52–S64. https://doi.org/10.1086/685710

Durand, A. (2019). *Toegang geweigerd: Digitale ongelijkheid in het slimme mobilite it st ijdper k*.

Engel, A. K., Friston, K. J., & Kragic, D. (Eds.). (2016). *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science*. The MIT Press. https://doi.org/10.7551/mitpress/9780262034326.001.0001

Fan, Z., Zhang, F., Loo, B. P. Y., & Ratti, C. (2023). Urban visual intelligence: Uncovering hidden city profiles with street view images. *Proceedings of the National Academy of Sciences*, *120*(27), e2220417120. https://doi.org/10.1073/pnas.2220417120

Fatmi, M. R., Chowdhury, S., & Habib, M. A. (2017). Life history-oriented residential location choice model: A stress-based two-tier panel modeling approach. *Transportation*

Research Part A: Policy and Practice, *104*, 293–307. https://doi.org/10.1016/j.tra.2017.06.006

Fields, C., & Levin, M. (2022). Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments. *Entropy*, *24*(6), 819. https://doi.org/10.3390/e24060819

Friston, K. (2018). *Of woodlice and men A Bayesian account of cognition, life and consciousness* (D. A. Friedman & M. Fortier, Interviewers) [Interview]. https://www.aliusresearch.org/uploads/9/1/6/0/91600416/friston_-_of_woodlice_and_men.pdf

Friston, K., Da Costa, L., Hafner, D., Hesp, C., & Parr, T. (2021). Sophisticated Inference. *Neural Computation*, *33*(3), 713–763. https://doi.org/10.1162/neco_a_01351

Friston, K., Da Costa, L., Sakthivadivel, D. A. R., Heins, C., Pavliotis, G. A., Ramstead, M., & Parr, T. (2023). Path integrals, particular kinds, and strange things. *Physics of Life Reviews*, *47*, 35–62. https://doi.org/10.1016/j.plrev.2023.08.016

Friston, K., Heins, C., Verbelen, T., Costa, L. D., Salvatori, T., Tschantz, A., Koudahl, M., Buckley, C., & Parr, T. (2024). *From pixels to planning: Scale-free active inference*.

Friston, K. J., Parr, T., Thomas Parr, de Vries, B., & de Vries, A. B. (2017). *The graphical brain: Belief propagation and active inference*. *1*(4), 381–414. https://doi.org/10.1162/netn_a_00018

Friston, K. J., Ramstead, M. J. D., Kiefer, A. B., Tschantz, A., Buckley, C. L., Albarracin, M., Pitliya, R. J., Heins, C., Klein, B., Millidge, B., Sakthivadivel, D. A. R., Smithe, T. S. C., Koudahl, M., Tremblay, S. E., Petersen, C., Fung, K., Fox, J. G., Swanson, S., Mapes, D., & René, G. (2022). *Designing Ecosystems of Intelligence from First Principles* (arXiv:2212.01354). arXiv. http://arxiv.org/abs/2212.01354

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: Active inference and agency. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00598

Friston, K., Sengupta, B., & Auletta, G. (2014). Cognitive Dynamics: From Attractors to Active Inference. *Proceedings of the IEEE*, *102*(4), 427–445. https://doi.org/10.1109/JPROC.2014.2306251

Garrido-Valenzuela, F., Cats, O., & Van Cranenburgh, S. (2023). Where are the people? Counting people in millions of street-level images to explore associations between people's urban density and urban characteristics. *Computers, Environment and Urban Systems*, *102*, 101971. https://doi.org/10.1016/j.compenvurbsys.2023.101971

Gemeente Zoetermeer. (2017). *Mobiliteitsvisie zoetermeer 2030*. https://www.zoetermeer.nl/_flysystem/media/mobiliteitsvisie-zoetermeer-2030.pdf

Geurs. (2018). *Transport Planning With Accessibility Indices in the Netherlands: Discussion Paper*.

Geurs, K. T., Haaijer, R., van Wee, B., & B Van Weeh. (2006). Option Value of Public Transport: Methodology for Measurement and Case Study for Regional Rail Links in the Netherlands. *Transport Reviews*, *26*(5), 613–643. https://doi.org/10.1080/01441640600655763

Geurs, K. T., & van Wee, B. (2004). Accessibility evaluation of land-use and transport strategies: Review and research directions. *Journal of Transport Geography*, *12*(2), 127–140. https://doi.org/10.1016/j.jtrangeo.2003.10.005

Geurs, K., & van Wee, B. (2023). *Accessibility: Perspectives, measures and applications*.

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2022a). *Spectral, Probabilistic, and Deep Metric Learning: Tutorial and Survey* (arXiv:2201.09267). arXiv. http://arxiv.org/abs/2201.09267

Ghojogh, B., Ghodsi, A., Karray, F., & Crowley, M. (2022b). *Spectral, Probabilistic, and Deep Metric Learning: Tutorial and Survey* (arXiv:2201.09267). arXiv. http://arxiv.org/abs/2201.09267

Gibson. (1986). Chapter 8 THE THEORY OF AFFORDANCES. In *The Ecological Approach to Visual Perception*.

Goekoop, R., & De Kleijn, R. (2021). How higher goals are constructed and collapse under stress: A hierarchical Bayesian control systems perspective. *Neuroscience & Biobehavioral Reviews*, *123*, 257–285. https://doi.org/10.1016/j.neubiorev.2020.12.021

Goekoop, R., & de Kleijn, R. (2023). *Hierarchical network structure as the source of hierarchical dynamics (power law frequency spectra) in living and non-living systems: How state-trait continua (body plans, personalities) emerge from first principles in biophysics* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2304.07094

Gong, F.-Y., Zeng, Z.-C., Zhang, F., Li, X., Ng, E., & Norford, L. K. (2018). Mapping sky, tree, and building view factors of street canyons in a high-density urban environment. *Building and Environment*, *134*, 155–167. https://doi.org/10.1016/j.buildenv.2018.02.042

Gontier, F. (2021). *Analysis and synthesis of urban sound scenes using deep learning techniques*.

Gramacki, P. (2021). *Unsupervised embedding and similarity detection of microregions using public transport schedules* (arXiv:2111.02405). arXiv. http://arxiv.org/abs/2111.02405

Gramacki, P., Leśniara, K., Raczycki, K., Woźniak, S., Przymus, M., & Szymański, P. (2023). SRAI: Towards Standardization of Geospatial AI. *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 43–52. https://doi.org/10.1145/3615886.3627740

Gramacki, P., Woźniak, S., & Szymański, P. (2021). *gtfs2vec—Learning GTFS Embeddings for comparing Public Transport Offer in Microregions*. https://doi.org/10.48550/ARXIV.2111.00960

Grover, A., & Leskovec, J. (2016). *node2vec: Scalable Feature Learning for Networks* (arXiv:1607.00653). arXiv. http://arxiv.org/abs/1607.00653

Gupta, J., Chen, Y., & Rammelt, C. (2024). Transport within earth system boundaries. *Npj Sustainable Mobility and Transport*, *1*(1), 5. https://doi.org/10.1038/s44333-024-00005-5

Haezendonck, E. (Ed.). (2007). *Transport project evaluation: Extending the social cost-benefit approach*. Elgar.

Hamilton, W. L., Ying, R., & Leskovec, J. (2018). *Inductive Representation Learning on Large Graphs* (arXiv:1706.02216). arXiv. http://arxiv.org/abs/1706.02216

Hansen, S. (1972). Utility, Accessibility and Entropy in Spatial Modelling. *The Swedish Journal of Economics*, *74*(1), 35. https://doi.org/10.2307/3439008

Harbers, A., van Amsterdam, H., & Spoon, M. (2022). *RUDIFUN 2022: Ruimtelijke Dichtheden en Functiemenging in Nederland*.

Hass, A. L., Runkle, J. D., & Sugg, M. M. (2021). The driving influences of human perception to extreme heat: A scoping review. *Environmental Research*, *197*, 111173. https://doi.org/10.1016/j.envres.2021.111173

He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition* (arXiv:1512.03385). arXiv. http://arxiv.org/abs/1512.03385

Higgs, C., Badland, H., Simons, K., Knibbs, L. D., & Giles-Corti, B. (2019). The Urban Liveability Index: Developing a policy-relevant urban liveability composite measure and evaluating associations with transport mode choice. *International Journal of Health Geographics*, *18*(1), 14. https://doi.org/10.1186/s12942-019-0178-8

Hillier, B. (2004). *Space is the machine*.

Hipolito, I., & Khanduja, A. (2024). *The Wicked and the Complex: A New Paradigm for Societal Problem-Solving*. https://doi.org/10.20944/preprints202407.1438.v2

Hipólito, I., & Podosky, P. (2024). *Beyond Control: Will to Power in AI*. https://doi.org/10.31234/osf.io/gn7ur

Hoffer, E., & Ailon, N. (2015). *Deep metric learning using Triplet network* (arXiv:1412.6622). arXiv. http://arxiv.org/abs/1412.6622

Hoogeboom, E., Peters, J. W. T., Cohen, T. S., & Welling, M. (2018). *HexaConv* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1803.02108

Hu, S., Gao, S., Luo, W., Wu, L., Li, T., Xu, Y., & Zhang, Z. (2023). Revealing intra-urban hierarchical spatial structure through representation learning by combining road network abstraction model and taxi trajectory data. *Annals of GIS*, *29*(4), 499–516. https://doi.org/10.1080/19475683.2023.2241526

Hu, S., Gao, S., Wu, L., Xu, Y., Zhang, Z., Cui, H., & Gong, X. (2021). Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Computers, Environment and Urban Systems*, *87*, 101619. https://doi.org/10.1016/j.compenvurbsys.2021.101619

Huang, T., Wang, Z., Sheng, H., Ng, A. Y., & Rajagopal, R. (2021). *Learning Neighborhood Representation from Multi-Modal Multi-Graph: Image, Text, Mobility Graph and Beyond* (arXiv:2105.02489). arXiv. http://arxiv.org/abs/2105.02489

Huh, M., Cheung, B., Wang, T., & Isola, P. (2024). *The Platonic Representation Hypothesis* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2405.07987

Huibregtse, O. (2021). *Kennisinstituut voor Mobiliteitsbeleid*.

Hutto, D. D., & Hipólito, I. (2021). *Updating Our Theories of Perceiving: From Predictive Processing to Radical Enactivism*.

Jana, A. (Ed.). (2021). *Advances in Urban Planning in Developing Nations: Data Analytics and Technology* (1st ed.). Routledge India. https://doi.org/10.4324/9781003091370

Jean, N., Wang, S., Samar, A., Azzari, G., Lobell, D., & Ermon, S. (2018). *Tile2Vec: Unsupervised representation learning for spatially distributed data* (arXiv:1805.02855). arXiv. http://arxiv.org/abs/1805.02855

Kaaronen, R. O., & Rietveld, E. (2021). Practical lessons for creating affordance-based interventions for sustainable behavior change. *One Earth*, *4*(10), 1412–1424. https://doi.org/10.1016/j.oneear.2021.09.013

Kaaronen, R. O., & Strelkovskii, N. (2020). Cultural Evolution of Sustainable Behaviors: Pro-environmental Tipping Points in an Agent-Based Model. *One Earth*, *2*(1), 85–97. https://doi.org/10.1016/j.oneear.2020.01.003

Kahneman, D. (2011). *Thinking, fast and slow*.

Kim, N., & Yoon, Y. (2022). *Effective Urban Region Representation Learning Using Heterogeneous Urban Graph Attention Network (HUGAT)* (arXiv:2202.09021). arXiv. http://arxiv.org/abs/2202.09021

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization* (Version 9). arXiv. https://doi.org/10.48550/ARXIV.1412.6980

Kipf, T. N., & Welling, M. (2017). *Semi-Supervised Classification with Graph Convolutional Networks* (arXiv:1609.02907). arXiv. http://arxiv.org/abs/1609.02907

Kmoch, A., Matsibora, O., Vasilyev, I., & Uuemaa, E. (2022). Applied open-source Discrete Global Grid Systems. *AGILE: GIScience Series*, *3*, 1–6. https://doi.org/10.5194/agile-giss-3-41-2022

Kmoch, A., Vasilyev, I., Virro, H., & Uuemaa, E. (2022). Area and shape distortions in open-source discrete global grid systems. *Big Earth Data*, *6*(3), 256–275. https://doi.org/10.1080/20964471.2022.2094926

Koudahl, M., van de Laar, T., & de Vries, B. (2023). *Realising Synthetic Active Inference Agents, Part I: Epistemic Objectives and Graphical Specification Language* (arXiv:2306.08014). arXiv. http://arxiv.org/abs/2306.08014

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

Kroesen, M. (2019). Residential self-selection and the reverse causation hypothesis: Assessing the endogeneity of stated reasons for residential choice. *Travel Behaviour and Society*, *16*, 108–117. https://doi.org/10.1016/j.tbs.2019.05.002

Leidelmeijer, K. (2004). *RIVM rapport 630950002 Kwaliteit van de Leefomgeving en Leefbaarheid*.

Leśniara, K., & Szymański, P. (2022). highway2vec: Representing OpenStreetMap microregions with respect to their road network characteristics. *Proceedings of the 5th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 18–29. https://doi.org/10.1145/3557918.3565865

Levering, A., Marcos, D., Van Vliet, J., & Tuia, D. (2023). Predicting the liveability of Dutch cities with aerial images and semantic intermediate concepts. *Remote Sensing of Environment*, *287*, 113454. https://doi.org/10.1016/j.rse.2023.113454

Levin, M. (2019). The Computational Boundary of a "Self": Developmental Bioelectricity Drives Multicellularity and Scale-Free Cognition. *Frontiers in Psychology*, *10*, 2688. https://doi.org/10.3389/fpsyg.2019.02688

Levin, M. (2022). Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds. *Frontiers in Systems Neuroscience*, *16*, 768201. https://doi.org/10.3389/fnsys.2022.768201

Li, Y., Huang, W., Cong, G., Wang, H., & Wang, Z. (2023). Urban Region Representation Learning with OpenStreetMap Building Footprints. *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1363–1373. https://doi.org/10.1145/3580305.3599538

Liang, Y., Zhu, J., Ye, W., & Gao, S. (2022). Region2Vec: Community Detection on Spatial Networks Using Graph Embedding with Node Attributes and Spatial Interactions. *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, 1–4. https://doi.org/10.1145/3557915.3560974

Litman. (2011). *Sustainability and Liveability*. Victoria Transport Policy Institute.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). *A ConvNet for the 2020s* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2201.03545

Luo, Y., Chung, F., & Chen, K. (2022). *Urban Region Profiling via A Multi-Graph Representation Learning Framework* (arXiv:2202.02074). arXiv. http://arxiv.org/abs/2202.02074

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

Marmot, Allen, Goldblatt, Boyce, Di McNeish, Grady, & Geddes. (2010). *Strategic review of health inequalities in England post-2010*.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370–396. https://doi.org/10.1037/h0054346

Mazzaglia, P., Verbelen, T., Çatal, O., & Dhoedt, B. (2022). The Free Energy Principle for Perception and Action: A Deep Learning Perspective. *Entropy*, *24*(2), 301. https://doi.org/10.3390/e24020301

Mazzaglia, P., Verbelen, T., & Dhoedt, B. (2021). *Contrastive Active Inference* (arXiv:2110.10083). arXiv. http://arxiv.org/abs/2110.10083

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (arXiv:1301.3781). arXiv. http://arxiv.org/abs/1301.3781

Miller, H. J., Witlox, F., & Tribby, C. P. (2013a). Developing context-sensitive livability indicators for transportation planning: A measurement framework. *Journal of Transport Geography*, *26*, 51–64. https://doi.org/10.1016/j.jtrangeo.2012.08.007

Miller, H. J., Witlox, F., & Tribby, C. P. (2013b). Developing context-sensitive livability indicators for transportation planning: A measurement framework. *Journal of Transport Geography*, *26*, 51–64. https://doi.org/10.1016/j.jtrangeo.2012.08.007

Millidge, B., Salvatori, T., Song, Y., Bogacz, R., & Lukasiewicz, T. (2022). *Predictive Coding: Towards a Future of Deep Learning beyond Backpropagation?* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2202.09467

Ministerie van Inrastructuur en Waterstaat. (2023). *Mobiliteitsvisie 2050 Hoofdlijnennotitie*.

Mink, R. (2023). *Investigating accessibility and related inequalities using a logsum metric grounded on a microscopic transport demand model*.

Moore, J. (2011). Behaviorism. *The Psychological Record*, *61*(3), 449–463. https://doi.org/10.1007/BF03395771

Moschella, L., Maiorca, V., Fumero, M., Norelli, A., Locatello, F., & Rodolà, E. (2023). *Relative representations enable zero-shot latent space communication* (arXiv:2209.15430). arXiv. http://arxiv.org/abs/2209.15430

Mouter, N., Koster, P., & Dekker, T. (2021). Contrasting the recommendations of participatory value evaluation and cost-benefit analysis in the context of urban mobility investments. *Transportation Research Part A: Policy and Practice*, *144*, 54–73. https://doi.org/10.1016/j.tra.2020.12.008

Nagatsu, M., Kaaronen, R. O., Salmela, M., & MacLeod, M. (2023). Cultural Niche Construction as a Framework for Reorienting Human−Environment Relations. *Topics in Cognitive Science*, *15*(3), 413–432. https://doi.org/10.1111/tops.12674

Norelli, A., Fumero, M., Maiorca, V., Moschella, L., Rodolà, E., & Locatello, F. (2023). *ASIF: Coupled Data Turns Unimodal Models to Multimodal Without Training* (arXiv:2210.01738). arXiv. http://arxiv.org/abs/2210.01738

Ororbia, A. G., & Friston, K. (2023). *Mortal Computation: A Foundation for Biomimetic Intelligence*. https://doi.org/10.31219/osf.io/epqkg

Oviedo, D., Sabogal, O., Villamizar Duarte, N., & Chong, A. Z. W. (2022). Perceived liveability, transport, and mental health: A story of overlying inequalities. *Journal of Transport & Health*, *27,* 101513. https://doi.org/10.1016/j.jth.2022.101513

Pacione, M. (1990). URBAN LIVEABILITY: A REVIEW. *Urban Geography*, *11*(1), 1–30. https://doi.org/10.2747/0272-3638.11.1.1

Pacione, M. (2003). Urban environmental quality and human wellbeing—A social geographical perspective. *Landscape and Urban Planning*, *65*(1–2), 19–30. https://doi.org/10.1016/S0169-2046(02)00234-7

Parr, T., Friston, K., & Zeidman, P. (2024). Active Data Selection and Information Seeking. *Algorithms*, *17*(3), 118. https://doi.org/10.3390/a17030118

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. The MIT Press.

Parvizi-Wayne, D., Sandved-Smith, L., Pitliya, R. J., Limanowski, J., Tufft, M. R. A., & Friston, K. (2023). *Forgetting Ourselves in Flow: An Active Inference Account of Flow States and How We Experience Ourselves Within Them*. https://doi.org/10.31234/osf.io/bqf8c

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. https://doi.org/10.3115/v1/D14-1162

Pereira, R. H. M., Schwanen, T., & Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, *37*(2), 170–191. https://doi.org/10.1080/01441647.2016.1257660

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 701–710. https://doi.org/10.1145/2623330.2623732

Pezzulo, G., Parr, T., Cisek, P., Clark, A., & Friston, K. (2024). Generating meaning: Active inference and the scope and limits of passive AI. *Trends in Cognitive Sciences*, *28*(2), 97–112. https://doi.org/10.1016/j.tics.2023.10.002

Porzi, L., Rota Bulò, S., Lepri, B., & Ricci, E. (2015). Predicting and Understanding Urban Perception with Convolutional Neural Networks. *Proceedings of the 23rd ACM International Conference on Multimedia*, 139–148. https://doi.org/10.1145/2733373.2806273

Raad voor de leefomgeving en infrastructuur. (2024). *Waardevol regeren: Sturen op brede welvaart*.

Raczycki, K. (2021). *Predicting the Location of Bicycle-sharing Stations using OpenStreetMap Data* (arXiv:2111.01722). arXiv. http://arxiv.org/abs/2111.01722

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., & Sutskever, I. (2021). *Learning Transferable Visual Models From Natural Language Supervision* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2103.00020

Rahmjoo, A., & Albarracin, M. (2023). *Intra-Active Inference I: Fundamentals*. https://doi.org/10.20944/preprints202306.1322.v1

Ramstead, M. J., Albarracin, M., Kiefer, A., Klein, B., Fields, C., Friston, K., & Safron, A. (2023). *The inner screen model of consciousness: Applying the free energy principle directly to the study of conscious experience* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/6afs3

Ramstead, M. J. D., Constant, A., Badcock, P. B., & Friston, K. J. (2019). Variational ecology and the physics of sentient systems. *Physics of Life Reviews*, *31*, 188–205. https://doi.org/10.1016/j.plrev.2018.12.002

Ramstead, M. J. D., Kirchhoff, M. D., & Friston, K. J. (2020). A tale of two densities: Active inference is enactive inference: *Adaptive Behavior*, *28*(4), 225–239. https://doi.org/10.1177/1059712319862774

Ramstead, M. J. D., Sakthivadivel, D. A. R., & Friston, K. J. (2024). *An approach to non-equilibrium statistical physics using variational Bayesian inference* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2406.11630

Ramstead, M. J. D., Sakthivadivel, D. A. R., Heins, C., Koudahl, M., Millidge, B., Da Costa, L., Klein, B., & Friston, K. J. (2023). On Bayesian mechanics: A physics of and by beliefs. *Interface Focus*, *13*(3), 20220029. https://doi.org/10.1098/rsfs.2022.0029

Reggiani, A., Bucci, P., & Russo, G. (2011). Accessibility and Impedance Forms: Empirical Applications to the German Commuting Network. *International Regional Science Review*, *34*(2), 230–252. https://doi.org/10.1177/0160017610387296

Reudink, M., Jeroen Bastiaanssen, Hans Elzenga, Ron Franken, Hans van Grinsven, Aldert Hanemaaijer, Christian Lennartz, Maarten van Schie, Johanna Schild, Kees Schotten, Roel van der Veen, Marijke Vonk, & Marlies Sanders. (2023). *Balans van de leefomgeving*.

Rietveld, E., & Kiverstein, J. (2014). A Rich Landscape of Affordances. *Ecological Psychology*, *26*(4), 325–352. https://doi.org/10.1080/10407413.2014.958035

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). *ImageNet Large Scale Visual Recognition Challenge* (arXiv:1409.0575). arXiv. http://arxiv.org/abs/1409.0575

Schwanen, T. (2020). Towards decolonial human subjects in research on transport. *Journal of Transport Geography*, *88*, 102849. https://doi.org/10.1016/j.jtrangeo.2020.102849

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Segundo-Ortin, M., & Raja, V. (2024). *Ecological Psychology* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009451413

Sims, M., & Pezzulo, G. (2021). Modelling ourselves: What the free energy principle reveals about our implicit notions of representation. *Synthese*, *199*(3–4), 7801–7833. https://doi.org/10.1007/s11229-021-03140-5

Smith, R., Varshney, L. R., Nagayama, S., Kazama, M., Kitagawa, T., & Ishikawa, Y. (2022). A computational neuroscience perspective on subjective wellbeing within the active inference framework. *International Journal of Wellbeing*, *12*(4), 102–131. https://doi.org/10.5502/ijw.v12i4.2659

Snellen, D., & Bastiaanssen, J. (2021). *Brede welvaart en mobiliteit*. Planbureau voor de Leefomgeving.

Ståhle, A. (2008). *Compact sprawl: Exploring public open space and contradictions in urban density*. School of Architecture, KTH.

Sun, F., Qi, J., Chang, Y., Fan, X., Karunasekera, S., & Tanin, E. (2023). *Urban Region Representation Learning with Attentive Fusion* (arXiv:2312.04606). arXiv. http://arxiv.org/abs/2312.04606

Sun, Y., Cheng, C., Zhang, Y., Zhang, C., Zheng, L., Wang, Z., & Wei, Y. (2020). *Circle Loss: A Unified Perspective of Pair Similarity Optimization* (arXiv:2002.10857). arXiv. http://arxiv.org/abs/2002.10857

Tan, How, Lee, & Zhiyi. (2024). *Building liveable & sustainable cities*.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

United Nations Department of Economic and Social Affairs. (2023). *The Sustainable Development Goals Report 2023: Special Edition*. United Nations. https://doi.org/10.18356/9789210024914

Vafaii, H., Galor, D., & Yates, J. L. (2024). *Poisson Variational Autoencoder* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2405.14473

Valk, A. van der, & Musterd, S. (1998). *Leefbare steden en een duurzame omgeving*. Van Gorcum.

van Cranenburgh, S., & Garrido-Valenzuela, F. (2023). *Computer vision-enriched discrete choice models, with an application to residential location choice*.

van de Laar, T., Koudahl, M., & de Vries, B. (2023). *Realising Synthetic Active Inference Agents, Part II: Variational Message Updates* (arXiv:2306.02733). arXiv. http://arxiv.org/abs/2306.02733

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

van Wee, B. (2016). Accessible accessibility research challenges. *Journal of Transport Geography*, *51*, 9–16. https://doi.org/10.1016/j.jtrangeo.2015.10.018

Vandenbroucke, D., Bucher, B., & Crompvoets, J. (Eds.). (2013). *Geographic Information Science at the Heart of Europe*. Springer International Publishing. https://doi.org/10.1007/978-3-319-00615-4

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (Version 7). arXiv. https://doi.org/10.48550/ARXIV.1706.03762

Veenhoven, R. (2000). *LEEFBAARHEID Betekenissen en meetmethoden*.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). *Graph Attention Networks* (arXiv:1710.10903). arXiv. http://arxiv.org/abs/1710.10903

Waltzman, R. (2017). *The Weaponization of Information: The Need for Cognitive Security*. RAND Corporation. https://doi.org/10.7249/CT473

Wang, J., & Biljecki, F. (2022). Unsupervised machine learning in urban studies: A systematic review of applications. *Cities*, *129*, 103925. https://doi.org/10.1016/j.cities.2022.103925

Wang, Z., Li, H., & Rajagopal, R. (2020). Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(01), 1013–1020. https://doi.org/10.1609/aaai.v34i01.5450

White, B., & Miller, M. (2024). Ambient smart environments: Affordances, allostasis, and wellbeing. *Synthese*, *204*(2), 48. https://doi.org/10.1007/s11229-024-04679-9

Wilke, G., & Portmann, E. (2016). Granular computing as a basis of human–data interaction: A cognitive cities use case. *Granular Computing*, *1*(3), 181–197. https://doi.org/10.1007/s41066-016-0015-4

Wilson, A. (1971). *Entropy in Urban and Regional Modelling (Routledge Revivals)* (0 ed.). Routledge. https://doi.org/10.4324/9780203142608

Wolsink, M. (1998). Leefbaarheid en duurzaamheid in de besluitvorming over infrastructuur. In *Leefbare steden en een duurzame omgeving* (pp. 138–150).

Wong, M., & Farooq, B. (2019). *Information processing constraints in travel behaviour modelling: A generative learning approach*. https://doi.org/10.48550/arXiv.1907.07036

Woźniak, S., & Szymański, P. (2021). Hex2vec—Context-Aware Embedding H3 Hexagons with OpenStreetMap Tags. *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 61–71. https://doi.org/10.1145/3486635.3491076

Wu, N., Zhao, X. W., Wang, J., & Pan, D. (2020). Learning Effective Road Network Representation with Hierarchical Graph Neural Networks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 6–14. https://doi.org/10.1145/3394486.3403043

Xiang, M. (2020). Region2vec: An Approach for Urban Land Use Detection by Fusing multiple Features. *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence*, 13–18. https://doi.org/10.1145/3404555.3404613

Xiao, L., Lo, S., Zhou, J., Liu, J., & Yang, L. (2021). Predicting vibrancy of metro station areas considering spatial relationships through graph convolutional neural networks: The case of Shenzhen, China. *Environment and Planning B: Urban Analytics and City Science*, *48*(8), 2363–2384. https://doi.org/10.1177/2399808320977866

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*, *180*, 148–160. https://doi.org/10.1016/j.landurbplan.2018.08.020

Zhang, M., Li, T., Li, Y., & Hui, P. (2020). Multi-View Joint Graph Representation Learning for Urban Region Embedding. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 4431–4437. https://doi.org/10.24963/ijcai.2020/611

Zhao, Y., Ke, Q., Korn, F., Qi, J., & Zhang, R. (2021). *HexCNN: A Framework for Native Hexagonal Convolutional Neural Networks* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2101.10897