Universität Zürich[UZH]

Geographisches Institut

A Computer Vision Framework for Automated Evaluation of School Route Safety in Zurich

GEO 511 Master's Thesis

Author: Claude Widmer, 20-707-279

Supervised by: Prof. Dr. Ross Purves, Dr. Tumasch Reichenbacher

Faculty representative: Prof. Dr. Ross Purves

22.10.2025

Abstract

This thesis develops and evaluates an automated computer vision framework to assess the safety of school routes in Zurich. Current safety maps rely on manual inspection and static data, which limits their regular update and scalability. To address this limitation, the study introduces a data-driven approach that combines street-level imagery, aerial orthophotos, and geospatial data with deep-learning-based detection and spatial analysis. Using YOLO-based object detection and depth estimation, safety-relevant infrastructure such as crossings, sidewalks and tram lines was identified and geolocated. These features were aggregated along the city's pedestrian network and evaluated through two complementary approaches: a rule-based scoring system and an machine-learning classifier. The resulting city-wide safety maps visualise relative safety levels for every street segment and reveal consistent spatial patterns across both methods. Peripheral and residential districts appear predominantly safe, while central transport hubs such as Bellevue and Central exhibit lower safety scores due to higher traffic complexity. Some limitations remain regarding computer vision accuracy, object localisation, and image coverage, especially under varying lighting or street conditions. The routing implementation successfully generated safer paths based on the computed safety scores. Overall, the developed framework represents a first practical test of a scalable, reproducible, and transparent approach to automated safety monitoring, supporting future research and evidence-based mobility planning towards safer school environments.

Acknowledgments

I would like to thank Prof. Dr. Ross Purves and Dr. Tumasch Reichenbacher for their supervision, constructive feedback, and ongoing support throughout this thesis. Their expertise and advice were highly valuable in shaping both the conceptual and methodological aspects of the work.

I am also grateful to Christian Schällibaum from the City Police Zurich, to Pascal Regli from Fussverkehr Schweiz, and to Andreas Reimers from the City of Zurich for their time, interest, and valuable feedback, which helped connect the analytical results with practical expertise from their respective fields.

Finally, I would like to thank my family and friends for their patience, motivation, and support throughout my studies, and especially my girlfriend Sarah for her support and encouragement throughout this time.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Background and Motivation

Urbanisation is one of the defining processes of the twenty-first century. More than half of the global population now lives in urban areas, and this proportion continues to rise. As cities expand and densify, they face increasing challenges in providing safe, inclusive, and sustainable living environments (United Nations 2019). This ambition is reflected in the United Nations Sustainable Development Goal 11, which calls for the development of "safe, resilient, and sustainable cities and communities" (United Nations 2025). Within this broader framework, mobility represents a central dimension of urban sustainability: how people move through the city directly affects quality of life, environmental performance, and social equity.

Creating sustainable cities requires mobility systems that are safe and accessible for all groups, including children. A sustainable urban environment cannot be achieved if children are treated merely as passengers rather than as active participants in urban mobility. Cities therefore need to provide walkable and safe environments that allow children to reach their destinations independently, supporting both sustainability and public health (Leden et al. 2014). Their ability to walk or cycle to school on their own reflects not only safety but also how supportive and inclusive a city's environment is.

Because of this, the safety of school routes has become an important topic in many cities, especially in Europe and Switzerland, where dense urban areas and mixed transport systems create both opportunities and challenges.

Many European cities, including Zurich, have launched initiatives to improve the safety of school routes and to reduce potential risks. In Zurich, the municipal police maintain an official School Route Safety Plan that marks recommended and critical routes (Zürich 2025). However, these maps are based on manual assessments and are updated only occasionally, making it difficult to capture ongoing changes in the urban environment. According to municipal sources, the City of Zurich is currently exploring new and more systematic ways of quantifying school route safety (Schällibaum, Pers. Comm.). This reflects a growing need for approaches that enable regular, objective, and scalable safety assessments.

To address this need, this thesis proposes an automated framework that relies primarily on street-level imagery while reducing dependence on predefined GIS datasets. Limiting the use of standardised GIS data allows the framework to draw directly on visual evidence of urban conditions, such as crosswalks, sidewalks, and signage, rather than on potentially incomplete or outdated geodata. Methodologically, the research adopts a case study approach focusing on the City of Zurich to examine the feasibility of such a computer vision–based framework and to identify suitable methods for its

implementation and evaluation.

Using computer vision techniques, relevant visual features can be systematically identified and evaluated across large areas, offering a scalable and reproducible alternative to manual assessments. This forms the conceptual basis for the framework introduced in the following section.

1.2 Research Objectives and Questions

The aim of this thesis is to design and evaluate an automated computer-vision–based framework to systematically assess the safety of school routes in the City of Zurich. The framework combines spatial and visual data to identify infrastructure features that are relevant to pedestrian safety and to translate them into a standardized, city-wide assessment.

To operationalize this aim, the research is structured around four guiding questions that address the key components of the framework:

Research Questions

1. **Feature Detection:** Which safety-relevant infrastructure elements can be reliably detected with computer vision, and how does performance vary across aerial and street-level imagery?

2. **Safety Scoring:** To what extent can machine-learning and rule-based scoring approaches provide reliable and interpretable assessments of school-route safety, and what are their respective strengths and limitations?

3. **Routing Integration:** How do safety-based routing models perform in generating feasible and safer school-route alternatives compared to conventional shortest-path routes?

4. **Validation:** How does the automated assessment compare to existing school route plans?

Together, these questions translate the overarching research aim into specific analytical steps. They ensure that the thesis not only demonstrates the technical feasibility of automated safety assessment but also evaluates its practical relevance, interpretability, and potential for application in broader urban contexts.

1.3 Research Focus and Framework Boundaries

The framework is developed and tested for the city of Zurich, taking into account Swiss traffic regulations, local infrastructure, and available spatial and visual data. Its focus lies on exploring how computer vision–based detection of safety-relevant infrastructure can be systematically integrated into network-based spatial analyses.

Rather than producing definitive safety ratings, the research evaluates the methodological feasibility and limitations of automated safety assessment, aiming to identify spatial patterns that may indicate potential safety concerns. These boundaries define the scope of the study and clarify that the framework represents an experimental, data-driven approach whose transferability to other cities is considered a potential, but secondary, outcome.

An important component of the framework is the integration with the existing street network. The detected safety-relevant features can be attached to road segments and subsequently used in network-based analyses such as routing algorithms. This allows not only for the evaluation of local safety conditions but also for the comparison of different routing strategies under varying safety assumptions.

The central focus is thus on exploring the potential and limitations of combining computer vision with network-based analysis and on understanding the methodological challenges of building such an automated framework. This exploration forms the core contribution of the thesis. Further methodological details and assumptions are presented in Chapter *Methodology*, and the limitations are discussed in Chapter *Discussion*.

1.4 Thesis Outline

This thesis is structured into seven chapters. Chapter *Introduction* introduces the topic, motivation, and research aims. Chapter *Theoretical Framework* reviews relevant literature on school-route safety and computational methods. Chapter *Study Area and Data* presents the study area and datasets, followed by Chapter *Methodology*, which details the analytical framework. Results are summarised in Chapter *Results*, and the findings are discussed and concluded in Chapters *Discussion* and *Conclusion*.

2 Theoretical Framework

This chapter outlines the theoretical foundations that guide the study of school-route safety and its operationalisation through data-driven and computer-vision-based methods. The chapter progresses from conceptual understandings of safety to the digital and computational frameworks through which safety is measured and interpreted.

2.1 Conceptualising School-Route Safety

This section synthesises findings from a broad literature base on school-route safety. It integrates behavioural, infrastructural, environmental, and institutional perspectives to show how safety is shaped, experienced, and managed in urban contexts. Instead of viewing safety as a purely technical measure, it is understood as a multidimensional concept that can be interpreted from different perspectives.

2.1.1 Children's Perception and Learning of Safety

Children's understanding of danger and their ability to act safely in traffic environments evolve through experience and social learning. Thomson et al. (2005) demonstrate that pedestrian competence depends on metacognitive skills such as anticipation, timing, and decision-making, which mature gradually through guided practice. These abilities enable children to judge traffic situations, predict vehicle movement, and coordinate their actions safely. Barton et al. (2006) find that structured instruction can improve crossing behaviour, but lasting effects occur only when adults model and discuss safety decisions. Without feedback and role models, children remain ill-prepared for unexpected situations. Morrongiello et al. (2007) highlight that gender and socialisation also influence risk-taking tendencies: boys are more likely to associate risk with excitement or competence, whereas girls are encouraged to act cautiously. These findings underline that safety is not inherent in the built environment. It is learned and continuously negotiated between children, adults, and space.

Perception and emotion further mediate these processes. Iqbal (2023) argue that urban safety must be understood as both physical and emotional: a child may feel unsafe on a technically secure path if it is dark, noisy, or deserted. The distinction between measured risk and felt safety is crucial for analysing school routes. While planners quantify risk through infrastructure metrics and accident statistics, families interpret safety through familiarity, visibility, and social trust. Kitchin (2017) therefore cautions that algorithmic representations of safety, such as automated risk maps, can only approximate lived experience. When integrated critically, however, they complement rather than replace human judgement by making large-scale safety patterns visible and comparable.

2.1.2 Spatial and Environmental Determinants

School-route safety is also shaped by its spatial and environmental context, which defines the physical conditions within which perception and behaviour take place. The configuration and quality of the built environment form its foundation: empirical studies consistently link accident risk to traffic volume, vehicle speed, and the continuity of pedestrian infrastructure. High traffic densities and missing sidewalks expose children to disproportionate risk (Rothman et al. 2015; Kumari 2021; Rahman et al. 2020), whereas slower traffic, speed enforcement, and well-maintained pedestrian routes reduce injuries and strengthen parental confidence (Rothman et al. 2019; Carlson et al. 2014; Mesfin et al. 2022).

The composition of the urban fabric further shapes exposure and perception. Mixed-use areas with shops and services encourage walking and provide social presence but can also generate higher traffic loads (Verhoeven et al. 2018; Milam et al. 2013; Wong et al. 2011; Zito et al. 2015). Environmental quality, including air and noise pollution, affects the comfort of walking to school. Unpleasant sensory conditions discourage active travel (Sarmiento et al. 2015; Peralta et al. 2020), whereas greenery, openness, and aesthetic maintenance increase children's comfort and autonomy (Dirks et al. 2018; Biernat et al. 2020). Urban environments that are legible, bright, and well-kept therefore not only reduce objective risk but also build trust in everyday mobility.

2.1.3 Infrastructure and Design Elements

Beyond environmental quality, the physical design of streets and crossings further shapes how safety is experienced and practised. Dedicated sidewalks and bicycle lanes separate vulnerable users from motorised traffic, reducing exposure and conflicts (Cieśla et al. 2022; Dyck et al. 2010; Mesfin et al. 2022). Rothman et al. (2013) and Carlson et al. (2014) show that such improvements directly increase rates of active commuting among children aged 10–15. Panter et al. (2010) and Feudjio Tezong et al. (2024) emphasise the value of traffic-calming measures such as narrower lanes, raised crossings, and curb extensions, which slow vehicles and improve visibility. Visibility-enhancing features such as lighting, clear sightlines, and passive surveillance are particularly important during early morning or late afternoon hours (Rahman et al. 2020; Milam et al. 2013; Ning et al. 2021). Where these elements are missing, parents are more reluctant to allow independent travel. Infrastructure thus functions not only as a protective system but also as a behavioural cue that signals whether walking and cycling are viable and trustworthy options.

2.1.4 Institutional and Community Dimensions

Beyond infrastructure, institutional frameworks and collective practices determine how safety is organised and sustained. The Safe Routes to School (SRTS) programme illustrates this integration of physical and behavioural interventions: infrastructure upgrades

combined with education campaigns effectively raise awareness and walking rates (McDonald et al. 2014; Greer et al. 2019; Carlson et al. 2014). Schwebel et al. (2014) emphasise that the most successful interventions pair environmental modifications with experiential training, reinforcing children's ability to apply safety knowledge in real contexts. At the regulatory level, speed limits and enforcement zones promote safer driving behaviour and strengthen a culture of responsibility in traffic environments (Rothman et al. 2015; D'Haese et al. 2011).

Parents and communities also mediate perceptions of safety. Oluyomi et al. (2014) and Vanwolleghem et al. (2016) find that parental fears—about traffic or social risks—strongly influence children's travel independence. Where trust in infrastructure is low, even objectively safe routes may remain unused (D'Haese et al. 2011; Mesfin et al. 2022). Community-based initiatives such as "walking school buses" combine social supervision with environmental improvement, reinforcing both real and perceived security (Dirks et al. 2018; Sarmiento et al. 2015; Abdullah et al. 2023). Osuret et al. (2022) demonstrate that participatory approaches must remain sensitive to local governance and culture, particularly in low- and middle-income contexts. Stakeholder dialogue is equally vital for automated approaches: Pascal Rengli et al. (2025) note that algorithmic safety assessments gain meaning only when interpreted alongside local expertise.

2.1.5 Integrated View and Empirical Synthesis

The reviewed literature demonstrates that school-route safety is shaped by interconnected physical, behavioural, and institutional factors. To make these relationships transparent, Table 2.1 summarises the main determinants identified across the research base, linking each key factor directly to its supporting evidence. This structure highlights recurring themes and the empirical consistency of findings across diverse contexts.

Table 2.1: Consolidated overview of key school-route safety factors and corresponding references.

Key Factor	**Main References**
Physical Infrastructure	
Sidewalk continuity, curb extensions, and pedestrian space design	(Rothman et al. 2013; Kim et al. 2020; Carlson et al. 2014)
Crosswalk visibility, signalisation, and surface design	(Greer et al. 2019; Panter et al. 2010)
Traffic calming (speed bumps, narrowed lanes, raised crossings)	(Rothman et al. 2015; Rahman et al. 2020)
Dedicated pedestrian or cyclist paths	(Cieśla et al. 2022; Mesfin et al. 2022)

Continued on next page

Table 2.1 – continued from previous page

Key Factor	**Main References**
Illumination, signage, and night-time visibility	(Milam et al. 2013; Rahman et al. 2020)
Built Environment Features	
Traffic volume and vehicle speed as predictors of injury risk	(Rothman et al. 2015; Kumari 2021)
Connectivity, accessibility, and intersection density	(Rothman et al. 2019; Feudjio Tezong et al. 2024)
Land use mix and destination diversity (residential–commercial balance)	(Verhoeven et al. 2018; Zito et al. 2015; Milam et al. 2013; Wong et al. 2011)
Air and noise pollution as deterrents for walking	(Sarmiento et al. 2015; Peralta et al. 2020)
Visual legibility, obstacles, and surface quality	(Biernat et al. 2020; Rahman et al. 2020; Mesfin et al. 2022)
Transportation Infrastructure	
Physical separation of motorised and non-motorised modes	(Cieśla et al. 2022; Dyck et al. 2010; Mesfin et al. 2022)
Protected crossings, refuge islands, and railing design	(Rothman et al. 2013; Mesfin et al. 2022)
Sightlines, lighting, and passive surveillance	(Rahman et al. 2020; Milam et al. 2013; Ning et al. 2021)
Network connectivity and safe multi-modal integration	(Peralta et al. 2020; Feudjio Tezong et al. 2024)
Policy and Institutional Factors	
Speed limit enforcement and school-zone regulations	(Rothman et al. 2015; D'Haese et al. 2011)
Safe Routes to School (SRTS) and combined education–infrastructure programmes	(McDonald et al. 2014; Carlson et al. 2014; Greer et al. 2019)
Pedestrian education and behavioural training	(Schwebel et al. 2014)
Legal enforcement mechanisms and institutional responsibility	(D'Haese et al. 2011; Greer et al. 2019)
Perception and Community Engagement	
Perceived traffic risk and fear of crime as behavioural barriers	(Oluyomi et al. 2014; Vanwolleghem et al. 2016)
Parental trust in infrastructure and independence decisions	(D'Haese et al. 2011; Mesfin et al. 2022)

Continued on next page

Table 2.1 – continued from previous page

Key Factor	**Main References**
Community participation and ownership in safety planning	(Abdullah et al. 2023; Sarmiento et al. 2015)
Cultural and governance context influencing interventions	(Osuret et al. 2022; Feudjio Tezong et al. 2024; Lizárraga et al. 2022)
Child-Centric Urban Design	
Age-appropriate crossings and spatial scale of design elements	(Kim et al. 2020; Zito et al. 2015)
Cognitive and physical development needs in mobility design	(Biernat et al. 2020; Zito et al. 2015)
Playfulness, sensory quality, and sense of comfort	(Peralta et al. 2020; Dirks et al. 2018)

2.1.6 Synthesis

In summary, school-route safety is not a fixed feature of infrastructure but a condition that emerges from social and spatial factors. The built environment creates the potential for safety, yet this potential becomes real only when supported by behaviour, institutions and public trust. According to Goodspeed (2020) and Naik et al. (2017), knowledge about safety in modern cities is increasingly shaped by data and algorithms. This development allows wider and more systematic observation but also raises questions about representation and possible bias. The following sections therefore explore how computational systems influence the ways in which safety is measured, visualised and addressed.

2.2 The Algorithmic City

Algorithmic urbanism refers to the increasing use of computational methods, optimisation, and artificial intelligence in processes of spatial decision-making. Son et al. (2023) defines it as a planning paradigm in which algorithms guide or co-produce decisions across scales, from building layouts to transport systems. Rather than replacing expertise, algorithmic planning formalises reasoning into models that simulate alternatives, optimise outcomes, and process large volumes of data. This development changes how planning knowledge is generated, emphasising modelling and prediction as central forms of reasoning.

Digital infrastructures and visual data archives have introduced new forms of urban visibility. Computer vision, sensors, and large-scale data platforms continuously record aspects of urban life, enabling highly detailed representations of the built environment. Naik et al. (2017) and Ding et al. (2021) demonstrate how machine learning applied to visual data can reveal structural patterns of urban form and transformation, thereby

producing new kinds of algorithmic visibility. These methods extend observation by translating urban phenomena into quantifiable features and categories, which in turn influence how evidence and spatial knowledge are constructed.

Algorithmic tools also support new modes of collaboration in planning. Data-driven models can make complex processes more transparent and comparable, thus providing a basis for dialogue among experts, stakeholders, and communities. Goodspeed (2020) suggests that scenario-based models may serve as instruments for shared learning rather than fixed prediction. In this sense, algorithmic systems can complement human interpretation when their assumptions and limitations are explicitly considered.

Philosophical perspectives further highlight the political dimensions of algorithmic governance. Lazar (2025) describes the algorithmic city as a form of governance in which computational systems participate in shaping how priorities are defined and justified. Algorithms are understood not merely as neutral instruments but as actors that contribute to determining what becomes visible, actionable, and legitimate within urban systems. This perspective underlines the entanglement of technology, perception, and authority in contemporary forms of urban management.

2.3 Computer Vision for Urban Safety

Computer vision has become an established field for analysing urban environments and built form. According to Starzyńska-Grześ et al. (2023), research in this domain focuses on two complementary directions: first, optimising algorithms for automated architectural and infrastructure feature recognition, and second, applying visual data to broader analytical questions linking spatial form with social meaning. Together, these strands position computer vision not only as a technical instrument but also as a lens for rethinking how cities are observed and interpreted.

A key development is the You Only Look Once (YOLO) family of models. Introduced by Redmon et al. (2016), YOLO reframed object detection as a single regression problem predicting bounding boxes and class probabilities at once. Jiang et al. (2022) trace its progress from YOLOv1 to YOLOv7, noting advances in network design, feature extraction, and loss functions. Due to its speed and simplicity, YOLO is widely used across domains such as autonomous driving and infrastructure monitoring. In safety research, its ability to process large image datasets in real time enables the automated identification of crossings, sidewalks, and other safety-relevant features.

Early work already illustrated the value of visual analytics for studying cities. Naik et al. (2017) used street-level imagery to predict physical urban change and to reveal social and environmental differences between neighbourhoods. Recent studies apply deep-learning detectors directly to traffic environments. Z. Yang et al. (2022) compared Mask R-CNN and YOLOv7 on Mapillary imagery and found that YOLOv7 achieved the best balance between precision (87.6%) and recall (72.6%). Kaya et al. (2023) report that YOLOv7 also outperformed Faster R-CNN in detecting pedestrian crosswalks, reaching

around 98% accuracy under variable conditions.

While object detection identifies visible features, it does not capture their spatial relationships or distance from the observer. Depth estimation addresses this limitation by adding a three-dimensional understanding of urban space. Advances in monocular depth estimation have improved three-dimensional understanding of urban space. Chen et al. (2019) integrated semantic segmentation into depth prediction, producing more coherent spatial representations and better alignment of features such as curbs and crossings.

Computer vision for urban analysis continues to advance rapidly. Each generation of models improves speed, generalisation, and interpretability, bringing automated city analysis closer to real-time application. As these methods mature, they become part of planning workflows, enabling evidence-based decisions while requiring attention to their technical and ethical constraints.

2.4 From Rule-Based Assessment to Machine Learning Scoring

While computer vision enables the automated observation of urban space, its analytical value depends on how visual detections are translated into meaningful indicators. The scoring of safety, which describes how risk is quantified and compared across space, represents a central step in this process. Two main paradigms shape this translation: rule-based assessment and machine learning based scoring. Each represents a distinct methodological logic for producing and validating urban knowledge.

Rule-based assessment represents the classical planning logic. It operates through explicit thresholds, weights, and penalties that link observable features such as sidewalks, crossings, or speed limits to safety categories. This approach reflects expert reasoning and normative judgement. It is transparent and reproducible, as every outcome can be traced back to a defined rule. In traffic safety research, such deterministic structures have long been used to explain crash severity or exposure risk (Savolainen et al. 2011; Silva et al. 2020). Yet, these systems simplify complex spatial relations. They assume stability and uniformity where in reality behaviour, context, and perception interact in unpredictable ways. Rule-based scoring provides interpretability but limited flexibility. It performs reliably within known contexts but struggles when applied to diverse or rapidly changing environments.

Machine learning follows a data-driven rather than a rule-based logic. Instead of prescribing relationships, models learn patterns directly from data. Algorithms such as Random Forests (Breiman 2001) or gradient boosting detect nonlinear interactions between features and observed outcomes. They generate probabilistic predictions that express uncertainty rather than fixed categories. In transport and urban safety studies, such models have achieved promising results in predicting crash risk and pedestrian safety (Iranitalab et al. 2017). Their strength lies in adaptability: as new data become available, the models update and refine themselves. This flexibility, however, reduces

transparency. The statistical learning process is not easily interpretable, and its internal reasoning may remain inaccessible to planners and policymakers.

Recent work has tried to bridge this gap between predictive power and interpretability. Post hoc explanation methods such as SHAP (SHapley Additive Explanations) decompose each prediction into additive feature contributions (Jain et al. 2020). This allows analysts to trace how much a variable such as street lighting or crossing density contributes to a safety score, even in complex ensembles. These tools aim to connect data driven inference with the transparency of rule based logic. They reflect a broader trend toward hybrid approaches in which empirical learning complements, but does not replace, expert knowledge and reasoning.

2.5 Research Gap

Previous research has greatly improved the understanding of school-route safety from behavioural, spatial and institutional perspectives. Many studies have identified how features of the built environment shape children's exposure and perception of risk. However, most of these studies rely on manual observation or surveys, which limits their scalability and makes it difficult to compare results across larger areas or time periods. Systematic and repeatable assessments of safety at city scale are still rare.

Recent progress in computer vision and data-driven urban analytics has opened new possibilities for analysing the built environment. Yet these techniques are rarely combined with planning-oriented frameworks or with approaches that explain how automated results relate to lived experience. Existing studies often focus on technical accuracy, while aspects such as transparency, interpretability and policy relevance receive less attention.

This thesis responds to these gaps by developing a computer-vision-based framework for the automated and interpretable assessment of school-route safety in Zurich. The framework links visual detection, spatial integration and explainable scoring to create a reproducible, transparent and transferable basis for analysing everyday mobility and safety.

3 Study Area and Data

3.1 Study Area

3.1.1 Geographical and Urban Context

The study area for this thesis is the municipality of Zurich (approx. 91.9 km^2). Located at the northwestern tip of Lake Zurich and traversed by the River Limmat, the city is administratively divided into 12 districts ("Kreise") and 34 neighbourhoods ("Quartiere"), further subdivided into more than 200 statistical zones. Official population statistics report approximately 430'000 residents at the end of 2024 (Statistik Stadt Zürich 2025).

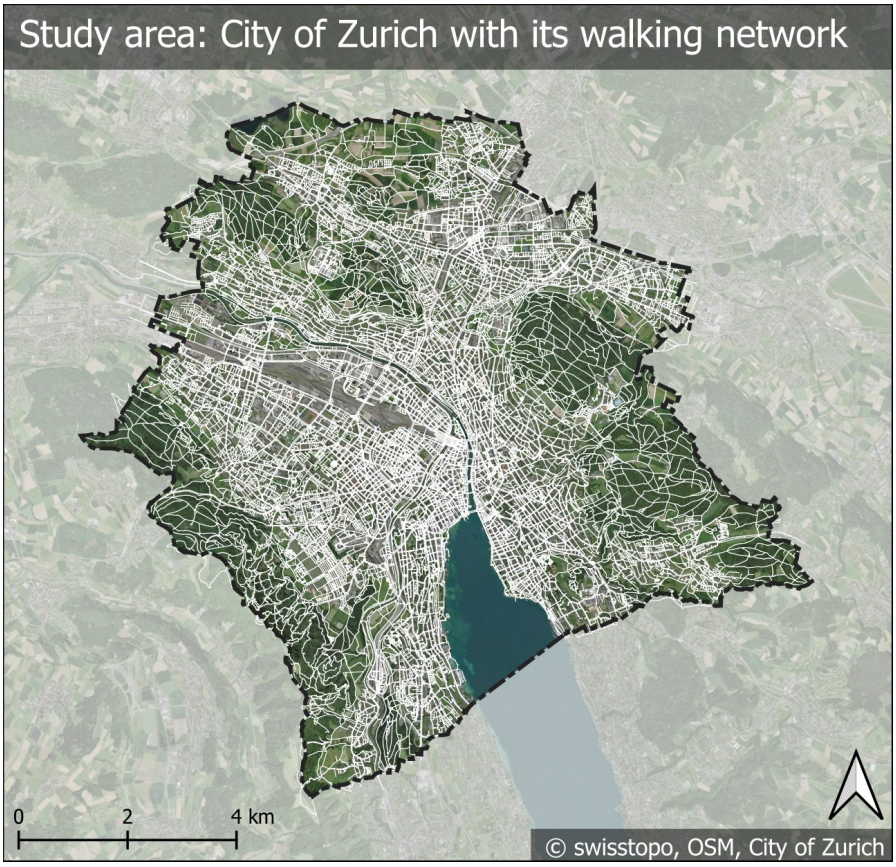


Figure 3.1: Study area: The City of Zurich with its pedestrian network.

3.1.2 Urban Infrastructure and Walkability

Zurich's compact layout, dense public transport network, and good local service coverage make walking a convenient option for daily trips (Klopper 2024).

At the same time, challenges remain. Studies point out that Greater Zurich still consumes land inefficiently, which weakens neighbourhood compactness and increases dependence on motorised transport (Wälty 2021). An analysis of so-called "10-Minute

Neighbourhoods" found that only one area in Zurich reaches a balanced mix of housing, workplaces and services within short walking distance (Wälty 2021). These findings connect to international discussions on neighbourhood sustainability, which highlight the importance of compact design, proximity and local accessibility (Khavarian-Garmsir et al. 2023).

On a European scale, Zurich performs relatively well. A recent comparative study ranked 121 metropolitan areas based on walking access to schools, shops, green areas and public services. Zurich was placed among the top twenty cities, showing that overall accessibility is high, even if some neighbourhoods still face gaps in connectivity and local service provision (Bartzokas-Tsiompras et al. 2023).

3.1.3 School Route Safety in Zurich and Switzerland

In Switzerland, school route safety (*Schulwegsicherheit*) has long been regarded as a public responsibility and is addressed through both research and local practice. Studies show that children today move around less independently than in the past, partly because parents worry about traffic dangers and other risks on the way to school (Rothman et al. 2015; Oluyomi et al. 2014). Yet research also finds that walking to school without adult supervision can strengthen children's confidence, independence, and perception of safety (Herrador-Colmenero et al. 2017).

From a legal perspective, the Swiss Federal Constitution guarantees every child access to adequate and free basic education according to Art. 19 and Art. 62 Abs. 2 BV (Bundesverfassung der Schweizerischen Eidgenossenschaft 2021). This implies that the journey to school must be reasonable and safely accessible. Swiss governance structures translate these legal obligations into local practice. The decentralised education system gives municipalities and schools significant autonomy in defining and implementing safety measures (Presence Switzerland 2024). In Zurich, the municipal police (*Stadtpolizei Zürich*) maintain an up-to-date GIS-based School Route Safety Plan (*Schulwegsicherheitsplan*), developed in collaboration with schools and parents (see Figure 3.2). This plan identifies recommended routes, marks crossing points, and highlights areas requiring special caution, thus linking official planning instruments with everyday guidance for families (Stadtpolizei Zürich 2025). It is continuously expanded as the city evolves and serves as a basis for targeted interventions such as speed reductions, crossing improvements, and awareness campaigns.

Beyond Zurich, cantonal authorities, safety organisations, and civil society groups also contribute to improving school route safety. National initiatives such as *Fussverkehr Schweiz* and *Safe2School* run public awareness campaigns and provide materials for schools and parents, promoting attention and caution in everyday traffic situations (Fussverkehr Schweiz 2025; Safe2School 2025). These activities complement municipal planning and underline that school route safety in Switzerland is not only a technical or infrastructural task, but also a shared social commitment.

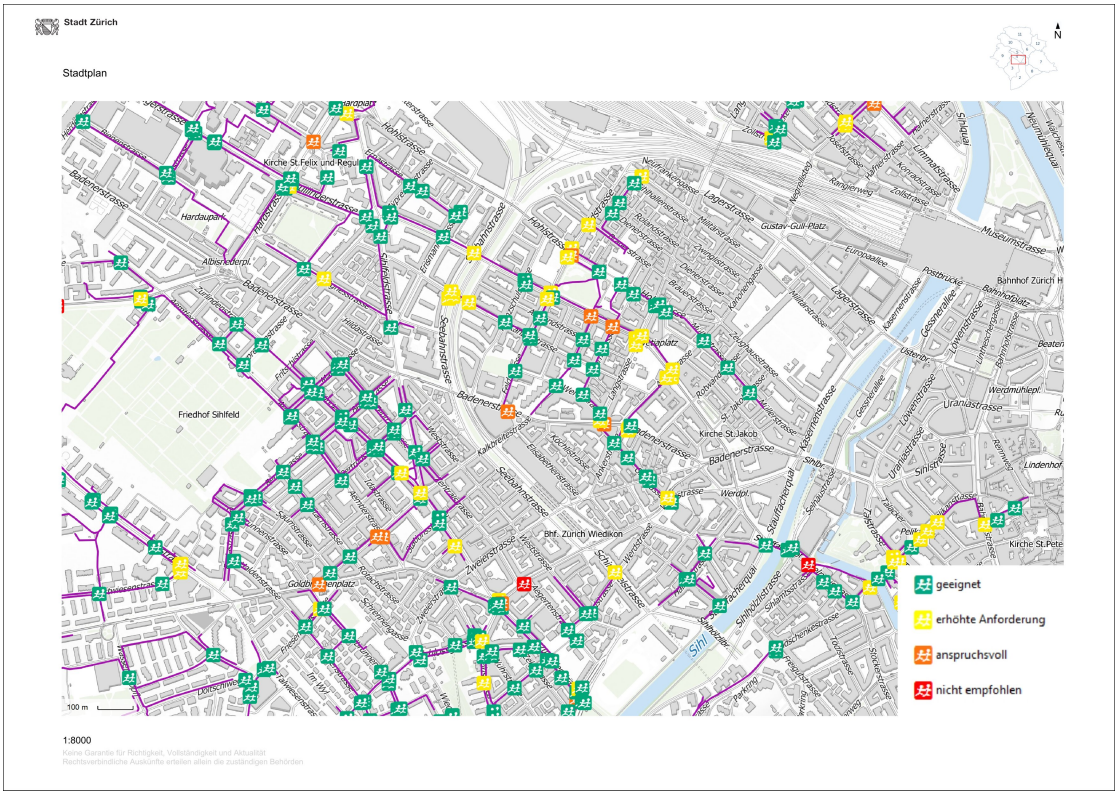


Figure 3.2: Excerpt from the official school route safety plan of the City of Zurich (Stadt Zürich 2025). Categories shown are recommended routes (*geeignet*), routes with increased requirements (*erhöhte Anforderung*), demanding routes (*anspruchsvoll*), and routes not recommended (*nicht empfohlen*).

3.2 Data

This thesis relies on a set of official municipal and federal geodata, complemented by open-source street network data and widely used benchmark datasets from computer vision. In addition, pretrained machine learning models were employed to enable depth estimation and object detection tasks. The focus is on primary data sources, which form the foundation for all subsequent analyses. Table 3.1 provides an overview of the datasets.

Table 3.1: Overview of data sources used in this thesis.

Use	**Parent data set / model**	**Attributes used**	**Source**
Street and pathway network	OSM extract (Geofabrik)	Geometry, highway type, sidewalks, crossings	OpenStreetMap / Geofabrik
Pedestrian and bicycle network	Official footpath and cycle network	Geometry, classification of foot/cycle paths	City of Zurich (GIS-Open-Data)
Aerial imagery	SWISSIMAGE orthophotos	High-resolution imagery	Swisstopo
Administrative data	AV-Daten	Parcels, building footprints, boundaries	Canton / City of Zurich
Administrative boundaries	SwissBoundaries3D	Municipal and cantonal boundaries, national borders	Swisstopo
Topographic road reference	SwissTLM3D (Transport layer)	Road geometries, object types (“10m”, “4m”, “6m”, “8m Strasse”)	Swisstopo
School route reference	Zurich School Route Plans	Suggested safe school routes	City of Zurich / Zurich Police
Semantic segmentation	Cityscapes; Mapillary Vistas	Training images, pixel-level labels	Cityscapes consortium; Meta/Mapillary
Depth estimation	Pretrained models	Depth prediction features	Hugging Face model hub
Object detection	YOLO11 pretrained model	Oriented bounding box segmentation model	Ultralytics
Tram detection training data	OGP ÖV-Geodaten (public transport lines)	Tram lines, filtered and segmented for training geometries	Amt für Raumentwicklung, Canton Zurich

The following subsections summarize the primary data sources and pretrained models.

3.2.1 SWISSIMAGE

SWISSIMAGE orthophotos provide high-resolution aerial imagery that (i) serves as a basemap for visualization and (ii) is used as input imagery for object-detection experiments with YOLO11 and oriented bounding boxes (OBB). The good quality and

nationwide coverage makes SWISSIMAGE valuable for both cartographic purposes and GIS-based detection workflows (Federal Office of Topography swisstopo 2023b).

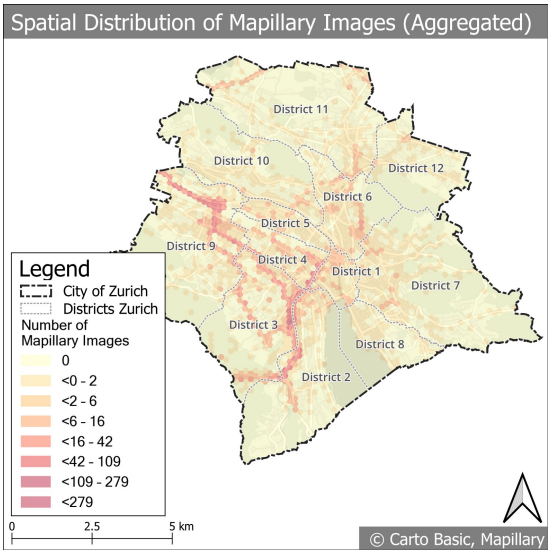
3.2.2 YOLO11

As outlined in the theoretical framework (see Section 2.3), the YOLO family of models enables efficient detection of urban infrastructure features relevant to safety analysis. Object detection and instance segmentation are performed using a pretrained *YOLO11* model. YOLO11 ("You Only Look Once", Version 11) combines high accuracy with efficient real-time performance. It supports polygonal masks (segmentation) and oriented bounding boxes (OBB), enabling robust detection of small and rotated objects. As a pretrained model, YOLO11 also constitutes part of the data basis used in this study, since it provides learned visual features derived from large-scale image datasets (Khanam et al. 2024; Ultralytics 2024).

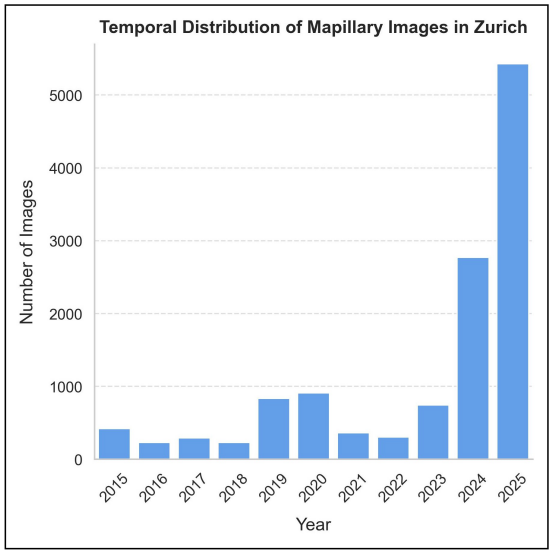
3.2.3 Mapillary

Mapillary is a large-scale, crowdsourced platform providing georeferenced street-level imagery collected from a variety of devices and users worldwide. The dataset covers different viewpoints, seasons, lighting, and weather conditions, resulting in heterogeneous image characteristics that reflect everyday urban environments. For the City of Zurich, Mapillary offers extensive visual coverage suitable for analysing mobility infrastructure and road environments.

Figure 3.3 shows the *spatial* and *temporal* distribution of available images for the City of Zurich (downloaded in June 2025). The spatial aggregation (left) indicates that most images cluster along major roads and in central districts, while coverage decreases toward residential areas and the city's periphery. This pattern is typical for crowdsourced data: frequently used streets are more likely to be documented, whereas less travelled areas remain underrepresented. The temporal distribution (right) illustrates a steady increase in contributions over time, with a marked growth between 2024 and 2025.



(a) Spatial distribution of Mapillary images in Zurich.



(b) Temporal distribution of Mapillary captures.

Figure 3.3: Spatial and temporal distribution of Mapillary images in Zurich (own analysis).

3.2.4 Mapillary Vistas

For semantic segmentation, the *Mapillary Vistas* dataset is used. It is a globally distributed, pixel-accurate benchmark with dense annotations for roads, sidewalks, traffic signs, pedestrians, and vehicles. The dataset's diversity supports generalization to heterogeneous urban conditions in Zurich (Neuhold et al. 2017). As shown in Figure 3.4, it provides labeling examples that demonstrate the variety and precision of annotated street-level imagery.

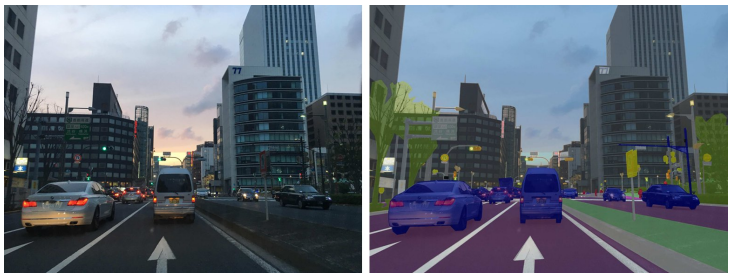


Figure 3.4: Qualitative labeling examples from the Mapillary Vistas dataset (adapted from Neuhold et al. (2017)).

3.2.5 Depth Estimation

For monocular depth estimation, the pretrained *Depth Anything V2 Small* model is used, available on the Hugging Face model hub (Hugging Face 2025). The model is based on a vision transformer and predicts dense depth maps for a wide range of real-world scenes. These depth maps are combined with object detections to analyse visibility, occlusions,

and the spatial relationships between sidewalks, roads, and street furniture (L. Yang et al. 2024).

3.2.6 Other datasources

In addition to the above, several further datasets are employed. Administrative context is provided by *SwissBoundaries3D*, which delivers harmonized municipal, cantonal, and national boundaries (Federal Office of Topography swisstopo 2023a). As a benchmark for validation, the official *Zurich School Route Plans* published by the City of Zurich and Zurich Police are included, representing recommended safe routes for schoolchildren (Stadt Zürich 2025). In addition, the official *Public Transport* dataset from the *Amt für Raumentwicklung Kanton Zürich* was used to extract tram line geometries (Amt für Raumentwicklung 2025). Finally, the street and pathway network is derived from two complementary sources: the official pedestrian and bicycle network of the City of Zurich (Stadt Zürich, Tiefbauamt 2024) and OpenStreetMap extracts obtained via Geofabrik (OpenStreetMap contributors 2025; Geofabrik GmbH 2025).

4 Methodology

4.1 Overview

The chapter Methodology introduces the methodological framework of the thesis. The overall workflow has been divided into four main phases, illustrated in Figure 4.1. Phase 1 involved preparing the pedestrian network and integrating additional data sources. Phase 2 involved feature extraction and geolocation of street-level and aerial imagery. Phase 3 classified school route safety according to the features extracted in Phase 2. Finally, Phase 4 covered the visualisation and presentation of the results. All code and scripts are available at `github.com/widmerc/Master_Thesis_Widmer`.

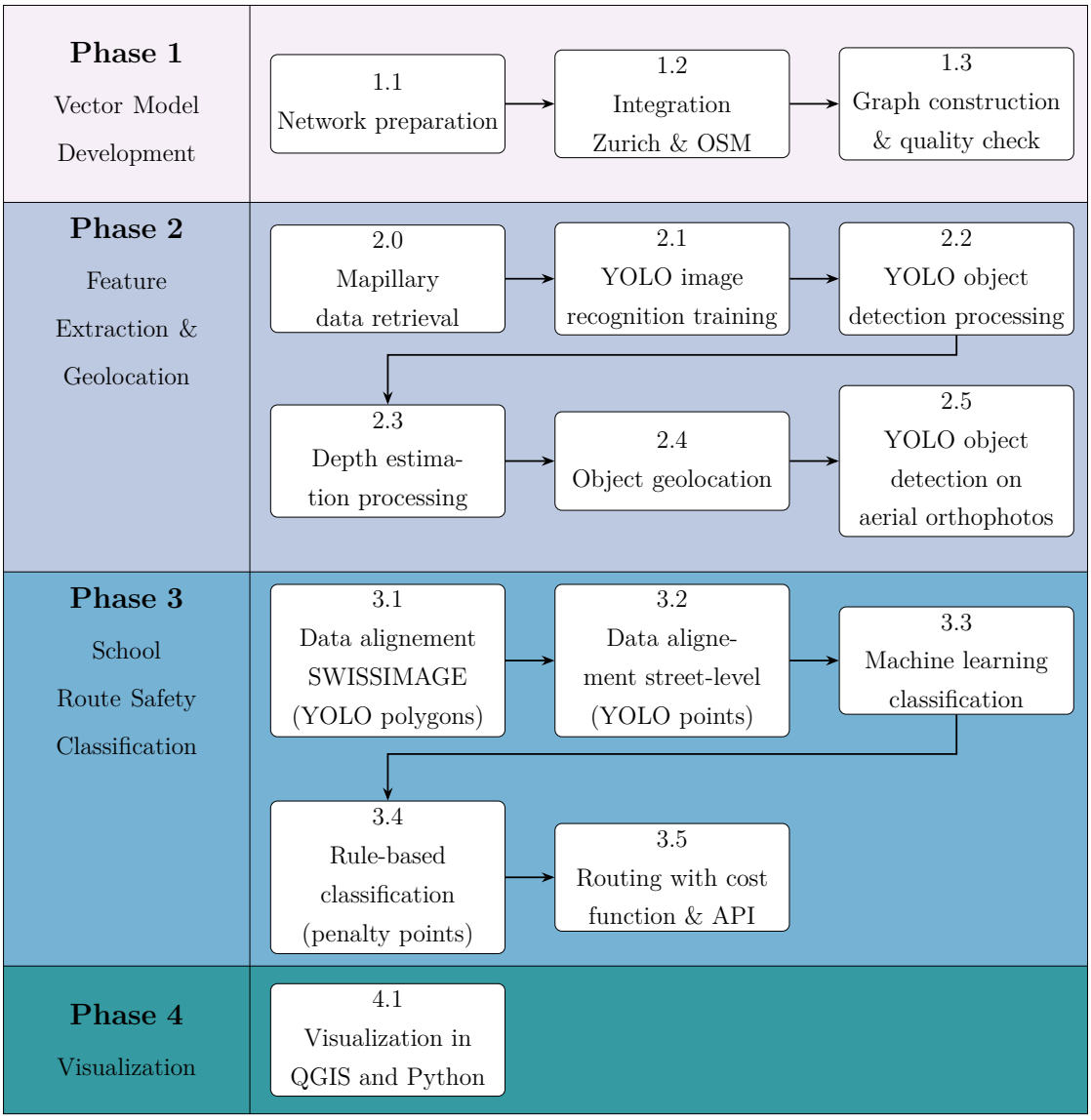


Figure 4.1: Overview of the methodology flowchart, illustrating the main phases and their substeps.

4.2 Phase 1: Vector Model Development

Phase 1 focused on creating a clear and routable pedestrian network model as the spatial foundation for all further analyses (see Figure 4.2).

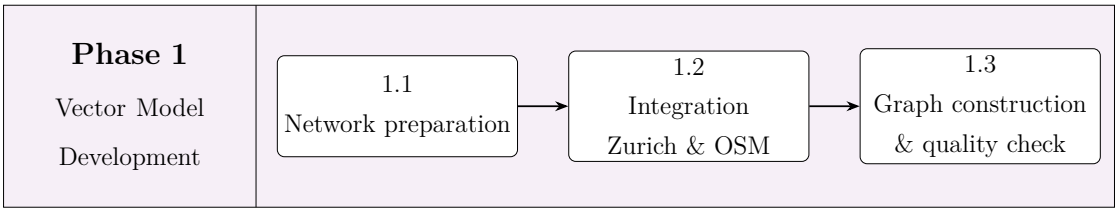
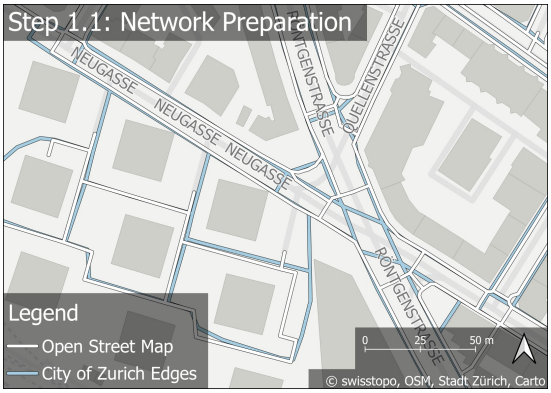


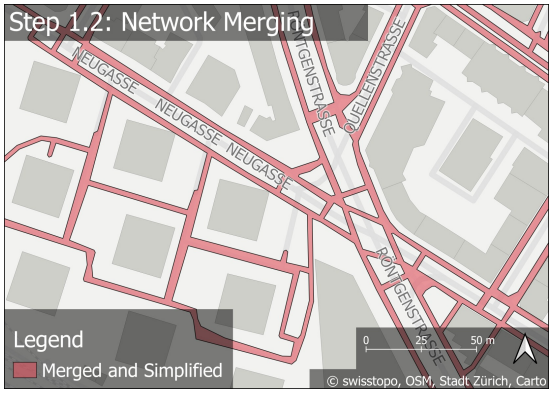
Figure 4.2: Workflow of Phase 1: Vector Model Development.

The workflow began with the extraction of pedestrian-relevant features from OpenStreetMap (OSM) (OpenStreetMap contributors 2025). From the wide range of available street types, only those suitable for walking were retained—such as `footway`, `pedestrian`, and `living_street`—while segments unsuitable for pedestrians, including `motorway`, `trunk`, or `service` roads restricted to vehicles, were excluded (see Figure 4.3a).

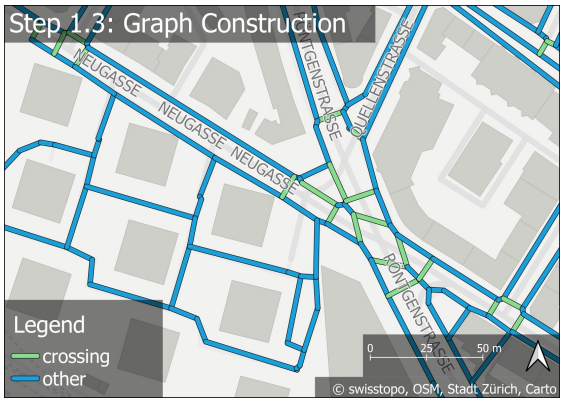
In order to ensure the completeness of the network, the OSM dataset was combined with the official *Pedestrian and Bicycle Network of the City of Zurich* (Tiefbauamt Stadt Zürich 2024). While OSM provides broad coverage, it often lacks detailed information on smaller paths, underpasses, or mixed-use areas such as shared spaces and residential streets. The official dataset helped fill these gaps and ensure that all pedestrian-accessible links were represented. Both datasets were spatially matched and aligned using a buffer-based approach, then merged and cleaned to create a continuous pedestrian surface (see Figure 4.3b). The merged geometries were then *skeletonized* into line features using GRASS GIS (GRASS Development Team 2024) within *QGIS* (QGIS Development Team 2025). Finally, the pedestrian lines were converted into a routable graph with `networkx` (Hagberg et al. 2008), including attributes such as segment length (`length_m`) and node connectivity. The SwissTLM dataset was loaded and filtered to retain relevant road object types (“10m Strasse”, “4m Strasse”, “6m Strasse”, and “8m Strasse”), which were then feature-matched with the pedestrian network to identify street connections and ensure correct transitions between pedestrian areas and the official road network (Federal Office of Topography swisstopo 2025). Quality checks ensured topological consistency and removed redundant or disconnected fragments (see Figure 4.3c).



(a) Step 1.1



(b) Step 1.2



(c) Step 1.3

Figure 4.3: Illustration of Phase 1: Network Model Development. Subfigures show (1.1) network preparation using OpenStreetMap and City of Zurich data, (1.2) merging and simplification of overlapping network segments and (1.3) graph construction with classification of sidewalks and crossings.

The final output of Phase 1 was a cleaned and fully routable pedestrian network. This graph was used as the structural foundation for linking computer-vision detections in Phase 2 and for the safety classification and routing analyses.

The entire workflow of Phase 1 was implemented in Python, combining `geopandas` (Van den Bossche 2022) and `shapely` (Gillies et al. 2025) for spatial processing, `networkx` (Hagberg et al. 2008) for graph construction, and `numpy` (Harris et al. 2020), `scipy` (Virtanen et al. 2020), `tqdm` (da Costa-Luis 2019) and `joblib` (The joblib developers 2025) for numerical computation, performance optimization, and process automation.

4.3 Phase 2: Computer Vision for Feature Extraction and Geolocation

In Phase 2, a computer vision framework was developed to automatically extract and spatially locate infrastructure features relevant to school-route safety (see Workflow Figure 4.4). Two different image sources were processed: (i) street-level photographs from the crowdsourced platform Mapillary and (ii) aerial orthophotos from the `SWISSIMAGE`

dataset. Combining these perspectives enabled the detection of features both from the ground and from above, providing a more complete representation of the pedestrian environment.

Custom `YOLO11` object detection models were trained for both datasets. The Mapillary model used annotated samples from the *Mapillary Vistas Dataset* (Neuhold et al. 2017) as ground truth, allowing the detection of crossings, sidewalks, and other safety-relevant elements from a pedestrian viewpoint. For the SWISSIMAGE data, a separate training set of more than 8,000 manually digitised and annotated features was created in QGIS, representing objects visible from an aerial perspective.

To further improve spatial accuracy, monocular depth estimation was applied to the detected objects in the Mapillary imagery. This allowed the approximate distance of each feature from the camera to be inferred and its real-world position to be linked to the pedestrian network. The integration of aerial and street-level detections resulted in a consistent, spatially referenced dataset that formed the basis for the subsequent safety classification of school routes.

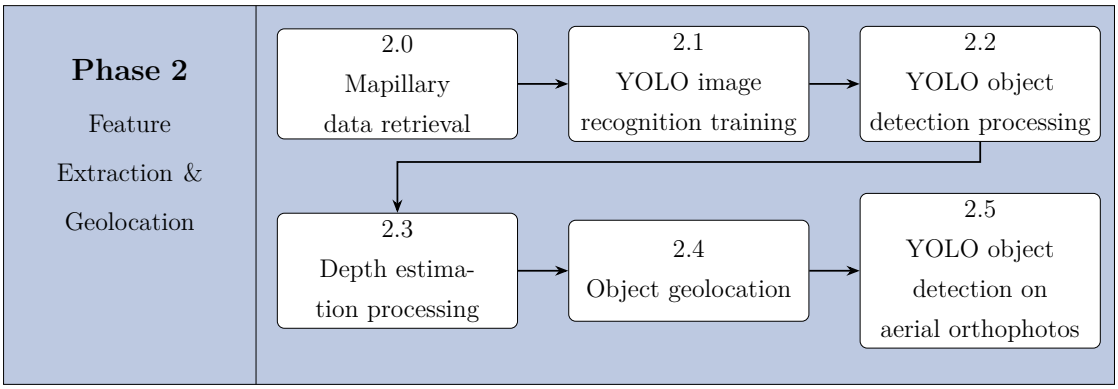


Figure 4.4: Workflow of Phase 2: Feature Extraction and Geolocation.

The implementation of this phase was carried out entirely in Python. It made use of the python libraries `torch` (Paszke et al. 2017), `YOLO11 ultralytics` (Jocher et al. 2024), `sahi` (Akyon et al. 2021; Akyon et al. 2022), `opencv` (Bradski 2000), `Pillow` (Clark 2015), and `transformers` (Wolf 2020) for object detection and depth estimation. Geospatial and data processing were handled with `geopandas` (Van den Bossche 2022), `shapely` (Gillies et al. 2025), `rasterio` (Gillies et al. 2013), `numpy` (Harris et al. 2020), `pandas` (The pandas development team 2020), `polars` (Vink et al. 2025), and `pyarrow` (Apache Arrow Developers 2025). For large-scale computation, `cupy` (Okuta et al. 2017), `numba` (Lam et al. 2015), and `scipy` (Virtanen et al. 2020) were used. Parallelisation and progress monitoring relied on `joblib` (The joblib developers 2025), `concurrent.futures`, `multiprocessing` and `tqdm` (da Costa-Luis 2019). Data retrieval was supported by the `mapillary SDK` (Beddow et al. 2021), `requests` (Chandra et al. 2015), `aiohttp` (Aio-Libs Community 2025), `aiofiles`, and `nest_asyncio`. Visualisation was carried out with `matplotlib` (Hunter 2007). Configuration and utilities included `yaml` (Simonov et al. 2025) and various Python standard modules (`os`, `sys`, `re`,

`math`, `time`, `gc`, `json`, `random`, `warnings`, `ast`, `unicodedata`, `pathlib`, `dataclasses`, `logging`).

A more detailed description of the six steps is provided in the following subsections.

4.3.1 2.0: Mapillary Data Retrieval

To analyse street-level imagery, Mapillary images covering the city of Zurich were first downloaded (see Figure 4.5). A bounding box was defined for the study area to ensure that all API requests remained within the same extent and could be reproduced.

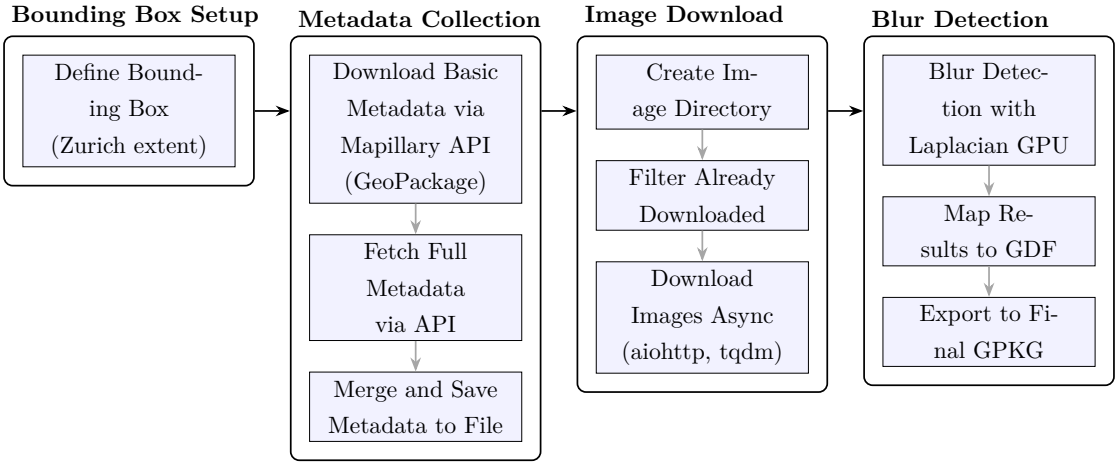


Figure 4.5: Workflow of the Mapillary data retrieval process, including bounding box setup, metadata collection, image download, and blur detection.

After defining the area, the Mapillary Python interface was used to fetch image metadata, including capture time, camera details, and download links. Because many images were involved, the requests were processed in parallel while respecting the Mapillary rate limits. After saving the metadata and image locations in a GeoPackage, the images were downloaded to a local storage drive on the workstation. To handle occasional API errors, the script was designed to resume automatically, skipping already downloaded files and continuing only with the missing ones.

Simply collecting the images was not enough, as their quality also had to be checked. For this reason a blur detection step was added. Instead of running on the CPU, the computation was moved to the GPU with `CuPy`, which made the process up to 500 times faster on large batches. The method used was the Laplacian variance: in principle, an image is converted to grayscale, convolved with a 3×3 Laplacian kernel, and the variance of the result is calculated (Bansal et al. 2016). The procedure can be summarized in three elements: (i) the kernel applied to the image, (ii) the variance formula, and (iii) the decision rule based on a threshold:

$$\text{(i)} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}, \qquad \text{(ii) } \sigma^2_{\nabla^2 I} = \mathrm{Var}(\nabla^2 I), \qquad \text{(iii) } \sigma^2_{\nabla^2 I} < \tau \;\Rightarrow\; \text{image is blurry,}$$

with a threshold of $\tau = 100.0$. A high variance indicates many edges and thus a sharp image, while a low variance indicates blur. All images below this value were flagged as blurry in the final GeoPackage (see Figure 4.6 for an example). The threshold value was chosen in line with similar implementations, where a value of 100 was found to reliably separate sharp from blurry frames in comparable imagery (Tian et al. 2026).

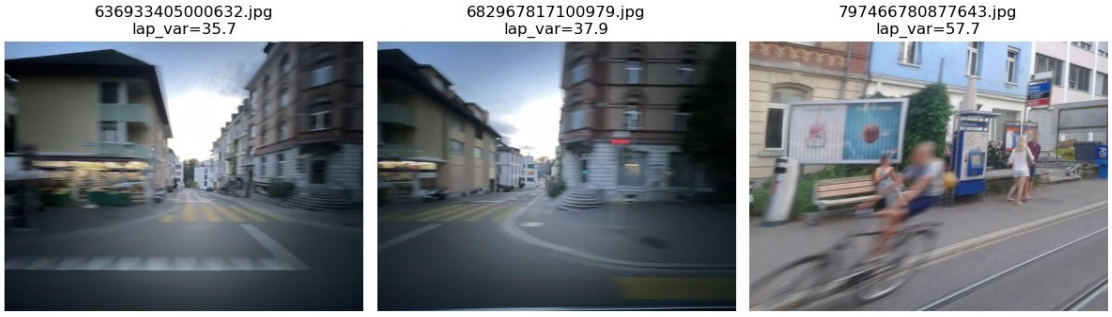


Figure 4.6: Example output of the blur detection step with unique Mapillary ID and calculated Laplacian variance (`lap_var`) (Mapillary 2025).

4.3.2 2.1: YOLO Image Recognition Training

After preparing the Mapillary images, the next step was to train a model for object detection and segmentation (see Figure 4.7 for the workflow).

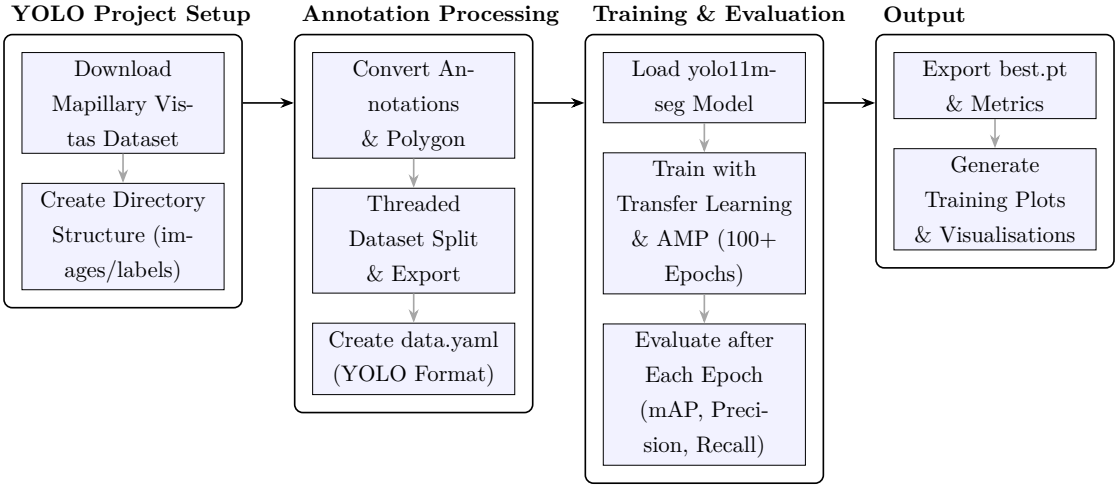


Figure 4.7: Workflow of YOLO image recognition training, covering dataset setup, annotation processing, model training and evaluation, and output generation.

For model training, the *Mapillary Vistas Dataset v2* was used, which contains more

than 18,000 street-level images with detailed polygon annotations across over 60 classes (Neuhold et al. 2017). The dataset was selected for its scale, annotation quality, and visual diversity. Since it is provided in COCO-style JSON format, the annotations were first converted into the YOLO-compatible text format. Each image received a corresponding `.txt` file containing the class index and polygon coordinates normalised to the image dimensions. The converted data were then organised in the standard YOLO directory structure (`images/train`, `labels/train`, etc.) together with a `data.yaml` file defining paths and class mappings for training.

With the dataset prepared, training was performed using Ultralytics YOLO11 segmentation variants (Redmon et al. 2016; Ultralytics 2024). YOLO was chosen for its favourable speed–accuracy trade-off and real-time inference capability, which is particularly useful for large datasets such as Mapillary (Jiang et al. 2022). Models were trained on an NVIDIA RTX 4080 (16 GB VRAM; 64 GB RAM).

YOLO models are available in several sizes, ranging from `n` (nano) and `s` (small) to `m` (medium), `l` (large), and `x` (extra-large). Smaller variants are optimized for higher inference speed and lower memory consumption, whereas larger variants provide higher accuracy at the cost of increased computational demand (Jiang et al. 2022).

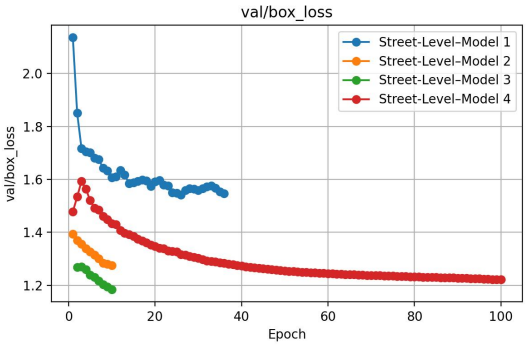
During experimentation, multiple model sizes and batch sizes were tested to evaluate performance and resource constraints. After several trials, four models were selected for extended training and subsequent comparison. The following four models were trained:

- **Street-Level–Model 1: YOLO11m-Early-Stopping (100 Epochs)**: medium size model trained for up to 100 epochs with early stopping (*patience* = 5); it stopped after 36 epochs.
- **Street-Level–Model 2: YOLO11m Fine-tuned from Model 1 (10 Epochs)**: first fine-tuning stage initialised from the best checkpoint of Model 1.
- **Street-Level–Model 3: YOLO11m Fine-tuned from Model 2 (10 Epochs)**: second fine-tuning stage initialised from Model 2.
- **Street-Level–Model 4: YOLO11s-Tuned (100 Epochs)**: small model trained for up to 100 epochs.

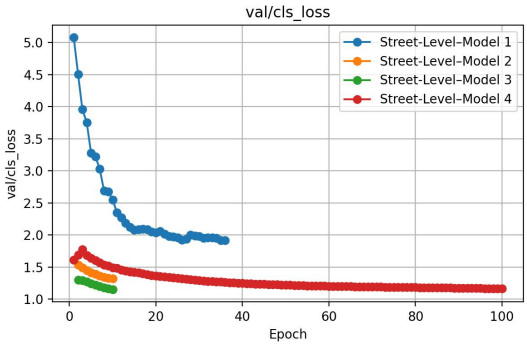
Validation was executed after every epoch and the best checkpoint was selected. For the model 1, early stopping was enabled (`patience`=5) to prevent overfitting and avoid unnecessary computation once validation metrics stabilized (Moraes et al. 2025). The training input size was set to `imgsz`=700 to maintain sufficient spatial detail for accurate object detection while ensuring feasible memory usage and training speed (Moraes et al. 2025). All other hyperparameters followed the Ultralytics defaults.

The learning curves in Figure 4.8 show consistent decreases in the *validation* losses. Relative to the early-stopped baseline (Street-Level–Model 1), the first fine-tuning stage (Street-Level–Model 2) already reduced losses markedly (e.g., *val/box_loss* from 1.39 to 1.27; baseline end: 1.55). The second stage (Street-Level–Model 3) further improved

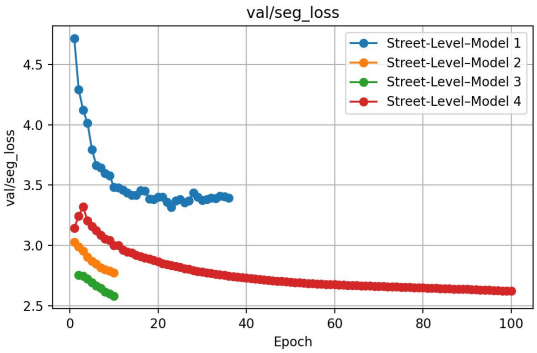
the medium model and yielded the lowest *val/box_loss* and *val/seg_loss* among all runs, while the small model (Street-Level–Model 4) achieved the lowest *val/cls_loss* after longer training. Based on these results, the second fine-tuned model (Street-Level–Model 3) was selected for the downstream detection step in Section 4.3.4.



(a) *val/box_loss*: bounding box regression



(b) *val/cls_loss*: class prediction



(c) *val/seg_loss*: segmentation masks

Figure 4.8: Validation loss curves for the four Street-Level runs (Street-Level–Model 1 to Street-Level–Model 4). Each plot shows the evolution of a specific validation loss during training.

All models were trained on subsets of the dataset, as training on all 18,000+ images would have exceeded the available computing time and hardware limits. The medium models (Street-Level–Model 1 to 3, `yolo11m-seg`) were trained on about 2,000 images each, while the small model (Street-Level–Model 4, `yolo11s-seg`) used around 10,000 images.

4.3.3 2.2: Depth Estimation Processing

Before running the object detection inference, it was necessary to generate depth information for the Street-Level images. This step aimed to move from a purely *2D* representation towards a *3D* understanding of the scenes. YOLO alone can recognise visible objects, but it does not provide information about their distance or spatial position relative to the camera (Redmon et al. 2016). It can only detect their location within the image and the geographic coordinates of where the photo was taken. This means that while objects can be identified and classified, their real-world depth and spatial

relationships remain unknown. Furthermore, the intrinsic parameters of the Mapillary cameras (such as focal length, optical centre, and lens distortion) are not consistently provided, which prevents the use of traditional stereo or structure-from-motion methods that rely on calibrated cameras or known baselines. Likewise, supervised depth networks depend on large labelled datasets with metric ground truth—conditions that are not met for Mapillary imagery (Gordon et al. 2019).

To address these limitations, the depth estimation followed the principle of *self-supervised monocular depth learning* as introduced by Gordon et al. (2019). Their approach showed that it is possible to estimate depth from uncalibrated image sequences by comparing how pixels change between neighbouring frames, without needing camera parameters or ground-truth depth data. Building on this idea, the *Depth-Anything V2* model (L. Yang et al. 2024) was used, which generalises the concept to large-scale single-image inference. The model predicts a depth value for every pixel directly from RGB images. It was trained on large and diverse image datasets without requiring camera calibration, which allows it to generalise well to new and unseen environments (L. Yang et al. 2024; Hugging Face 2025) This makes it well suited for heterogeneous sources like Mapillary, where camera information is often missing or incomplete. The produced depth maps give an approximate but consistent geometric view of each scene and serve as the foundation for later 3D analyses.

Since the workflow covered the entire city of Zurich, the input dataset comprised more than 1.2 million Mapillary images. Therefore, the processing pipeline had to be designed for robustness, efficiency, and scalability (see Figure 4.9). The workflow began with the list of images validated for blurriness in Step 2.0 (see Subsection 4.3.1). To prevent corrupted files from entering the pipeline, all `.jpg` images were checked using the Python Imaging Library (PIL). This parallelised validation ensured that millions of images could be verified within minutes, producing a clean and reliable input list for further processing.

Directly processing the original full-resolution Mapillary images would have been prohibitively slow, so a dedicated downscaling step was added. All valid images were resized to 50% of their original dimensions using bilinear interpolation. This reduced memory usage by roughly half and significantly increased throughput while maintaining sufficient structural detail for meaningful depth prediction. The script was fully parallelised, skipped already processed files, and produced a mirror directory of reduced images.

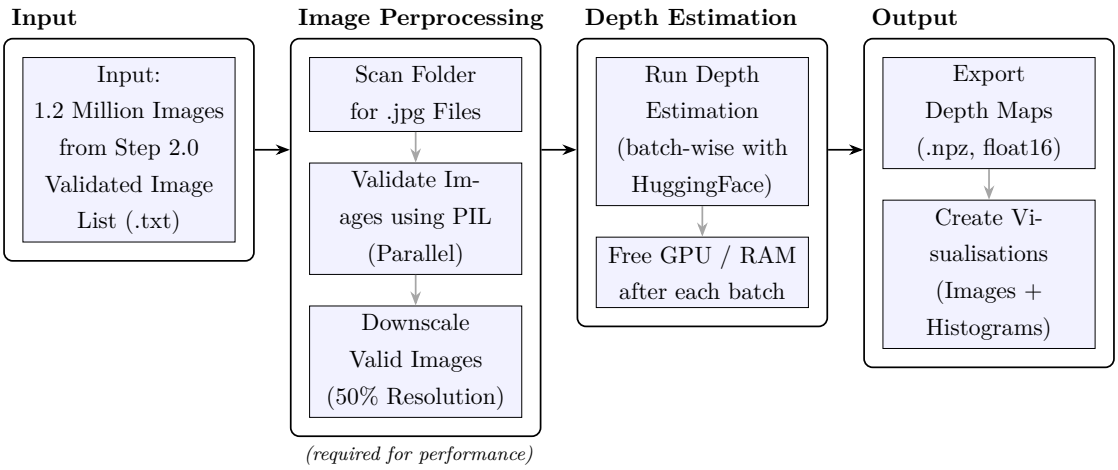


Figure 4.9: Workflow of the depth estimation pipeline, including image validation and preprocessing, depth estimation, and output generation.

Depth estimation was performed using a Hugging Face pipeline based on the `depth-anything-v2` model family (Hugging Face 2025; L. Yang et al. 2024). During development, two configurations were tested — a small and a large model variant. The large model produced slightly sharper object boundaries and smoother depth transitions, whereas the small model delivered much faster inference with only minor quality differences. Considering the dataset size and available hardware, the small model was chosen as the optimal compromise between accuracy and efficiency. Example outputs are presented in Results Section 5.3.

The model generated raw depth maps as floating-point arrays. To save storage space, these arrays were converted to `float16` precision and stored as compressed `.npz` files, reducing file size by about 75% compared to uncompressed `.npy`. The pipeline automatically skipped already processed files, allowing interrupted runs to resume smoothly and making the workflow robust against crashes.

The resulting depth maps formed the geometric basis for the subsequent detection phase, in which object inference was applied to the same set of street-level images.

4.3.4 2.3: YOLO Object Detection Processing

After generating the depth maps, object detection was performed on the same street-level images to identify safety-relevant infrastructure elements. Using the refined `Street-Level-Model 3` (see Section 4.3.2), relevant objects were automatically detected.

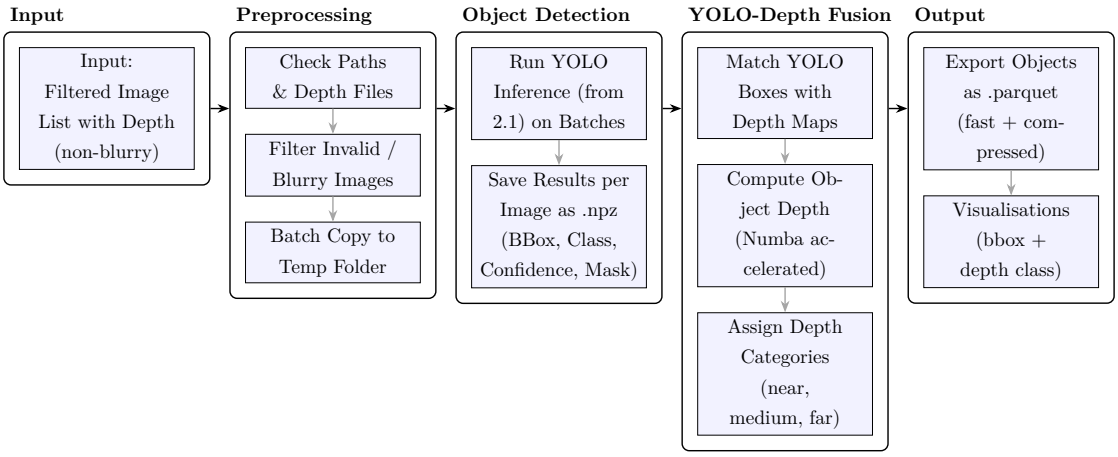


Figure 4.10: Workflow of YOLO object detection processing (Step 2.3), including preprocessing, YOLO inference, depth fusion, and dataset export.

Before running the inference, the list of valid images was cross-checked with the available depth maps to ensure consistency. Images flagged as blurry (see Section 4.3.1) or lacking depth results were excluded. The remaining files were divided into temporary batch folders to enable parallel processing, as running inference on the full dataset of more than 1.2 million images would have exceeded the available resources. The YOLO model was then used for inference, configured with a batch size of 16 (limited by the Nvidea RTX 4080 GPU's 16 GB VRAM), an input image size of 1280 px, and a confidence threshold of 0.25. These parameters were determined empirically to achieve an optimal balance between accuracy and runtime efficiency. The model generated segmentation masks, bounding boxes, confidence scores, and predicted classes for each image. All outputs were stored as `.npz` files. Figure 4.11 shows an example of the detection results.

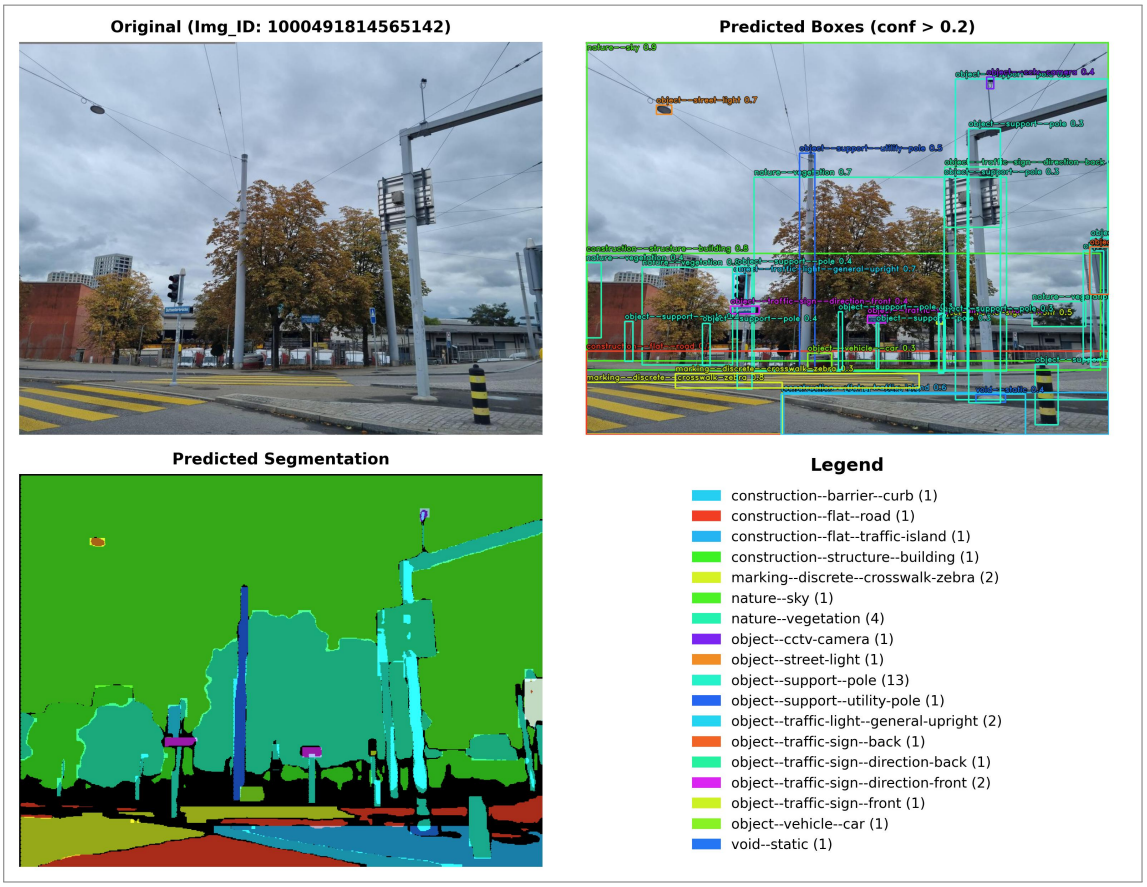


Figure 4.11: Example output of Street-Level–Model 3 showing input image, detected objects (confidence > 0.2), and segmentation masks.

Then, the YOLO detections were combined with the depth estimations that had been computed earlier (see Section 4.3.3). The relevant depth values for each bounding box were extracted from the saved .npz files. To ensure a robust estimate, the median depth of all valid pixels within the corresponding segmentation mask was calculated. Using the median made the method robust against outliers, such as holes in the depth map or small regions with erroneous values. For the calculations the python library `Numba` was used for optimization. This procedure effectively added a third dimension to the YOLO outputs.

Similar strategies have been reported in recent research combining YOLO-based segmentation and deep-learning-derived depth maps. Lin et al. (2024) demonstrated that integrating YOLOv8n-seg with monocular and stereo depth estimation models (e.g., MiDaS, Depth Anything, NeRF) enables accurate 3D distance estimation and spatial localization of detected objects. Their study highlights the effectiveness of median-based depth aggregation for robust distance prediction without requiring ground-truth depth data. Inspired by this approach, the present framework applies a comparable fusion method between YOLO detections and pixel-level depth maps to enrich each object with a third spatial dimension.

For interpretability, each detected feature was subsequently assigned to a qualitative

depth category (*very near*, *near*, *medium*, *far*) based on quantiles of the overall depth distribution. These classes were also used for visual quality control, where bounding boxes were colour-coded by distance to verify plausibility (e.g., a pedestrian in the foreground labelled as *very near*, a car further down the street as *far*). Figure 4.12 illustrates this fusion, with detected objects colour-coded according to their estimated distance class.

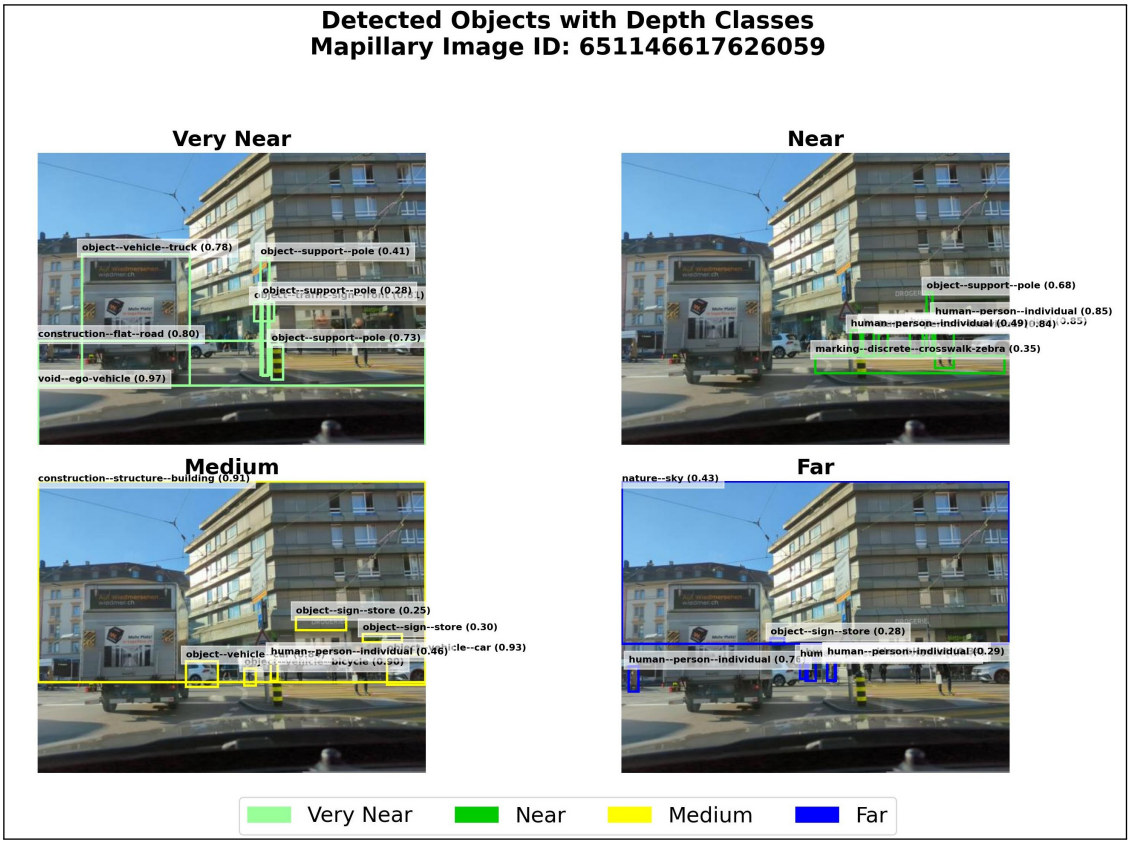


Figure 4.12: Example of YOLO + depth fusion. Detected objects are colour-coded by their assigned distance class (*very near*, *near*, *medium*, *far*).

A .parquet file was used to save all the detected features. Each record included the image ID, object category, confidence rating, bounding box coordinates, optional segmentation mask, median depth value, and the assigned distance category. The Parquet format was chosen for its high compression, efficient sequential reads, and seamless integration with analysis libraries such as `pandas` and `pyarrow` (Saeedan et al. 2022). This ensured that the storage and memory requirements for millions of object records remained manageable.

The result of this phase was a structured dataset containing each detected object with its class, position, and depth information, which served as the basis for geolocation in Section 2.4 Object Geolocation.

4.3.5 2.4 Object Geolocation

Building on this dataset, the next step projected the detected objects from image space into geographic coordinates. The workflow is illustrated in Figure 4.13. Because the

Mapillary images do not provide detailed camera calibration or reference depth data, a simplified geometric projection approach was used.

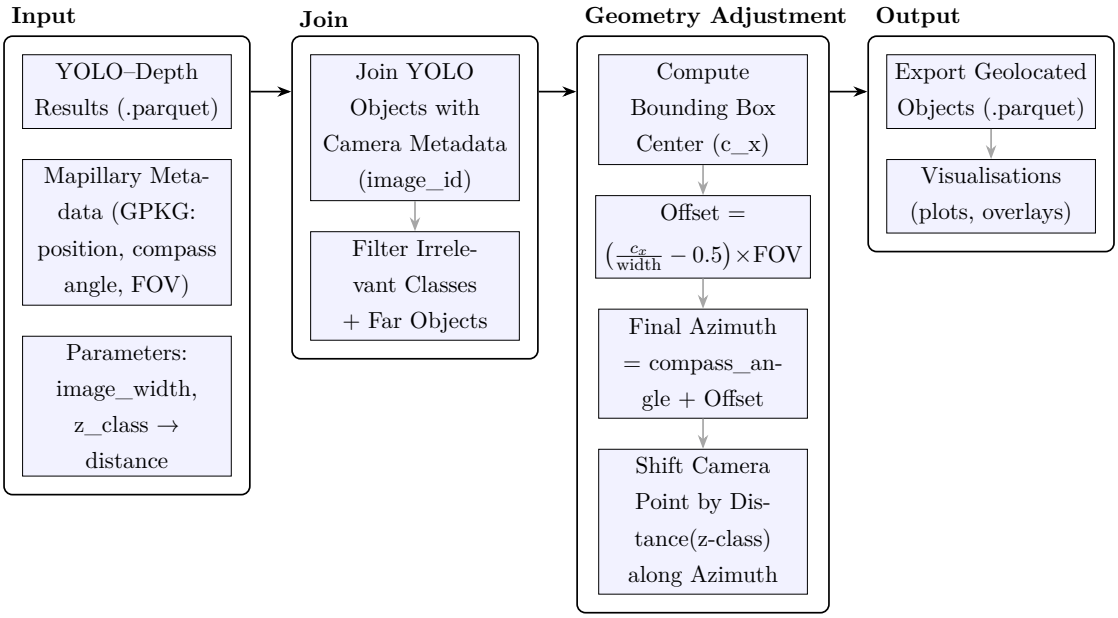


Figure 4.13: Workflow of object geolocation (Step 2.4), including join with camera metadata, geometry adjustment (offset, azimuth, shift), and output generation.

The inputs for the algorithm consisted of (i) the Parquet file with all YOLO+depth detections, (ii) the Mapillary metadata in GeoPackage format (image position, compass angle, field of view), and (iii) parameters such as the image width (1024 px), the camera field of view (90°), and class-based distance offsets derived from the `z_class`. The offsets were defined as 5 m for *very near*, 7.5 m for *near*, 20 m for *medium*, 50 m for *far*, and 100 m for *very far*. The chosen offset distances were determined empirically through iterative testing — by visually comparing shifted object positions with their appearance in the images until the placements appeared realistic. In practice, only objects in the categories *very near*, *near*, and *medium* were retained for geolocation, while more distant objects were excluded due to their limited positional reliability. Several irrelevant classes (e.g. sky, terrain, banners) were also removed.

First, the detection results were joined with the camera metadata using the `image_id`. This ensured that each detection was linked to its corresponding position, compass angle, and field of view.

In the geometry adjustment stage, image coordinates were transformed into azimuth and distance. The horizontal center c_x of each bounding box was used to determine how far the object lies from the central viewing direction. Since the Mapillary metadata provide only the horizontal field of view but no detailed camera calibration, the angular offset $\Delta\theta$ was approximated linearly from the relative horizontal image position. This simplified geometric adjustment follows the same principle as in previous work on street-level image geometry, where pixel coordinates are converted into azimuth or elevation angles based on known field-of-view parameters (Ning et al. 2022).

The underlying idea is illustrated in Figure 4.14: the center of the image corresponds to the optical axis (0°), while the left and right image edges represent $-\frac{1}{2}$ and $+\frac{1}{2}$ of the total field of view, respectively. This means that each pixel position in the image can be mapped proportionally to an angular deviation from the image center.

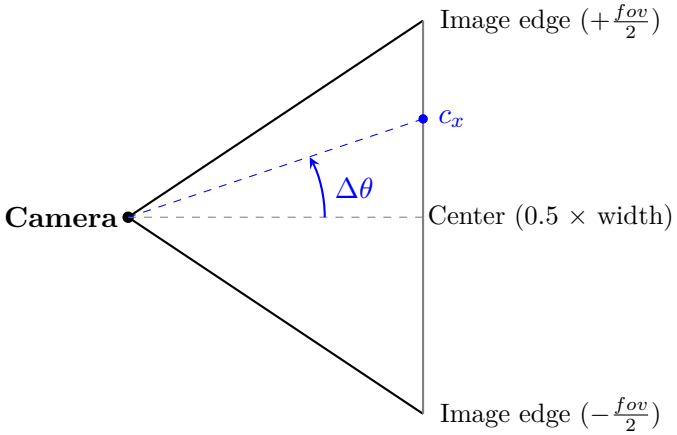


Figure 4.14: Mapping of the pixel position c_x to the horizontal viewing angle $\Delta\theta$.

Based on this geometric assumption (see Figure 4.14), the offset angle for each object was calculated as

$$\Delta\theta = \left(\frac{c_x}{\text{image\_width}} - 0.5\right) \times fov.$$

The final azimuth was then obtained by adding this offset to the camera compass angle. Starting from the camera position, the object location was derived by shifting the point along this azimuth by the distance associated with the object's `z_class`. All computations were carried out in the Swiss LV95 projection (EPSG:2056), ensuring that offsets were applied in metric units. Panoramic images were treated as a special case: here the camera position itself was retained, as no single viewing direction could be defined. The results were written to Parquet files in two stages: first as a joined dataset (original camera positions), and then as an adjusted dataset (shifted object positions).

A quantitative evaluation of the geolocation results, including displacement statistics, runtime, and visual examples, is presented in Chapter Results.

4.3.6 2.5: YOLO Object Detection on SWISSIMAGE

In addition to the street-level imagery, the object detection workflow was also applied to aerial data (see Figure 4.15). This was done because infrastructure elements such as tram tracks, crosswalks, and road markings are more easily detected from a nadir perspective, where their geometry is fully visible, as also shown by Antwi et al. (2024). For this step, the *SWISSIMAGE* dataset was used as a complementary source. The orthophotos have a ground resolution of 10 cm per pixel, which is detailed enough to capture small road features while also covering the whole country. The city of Zurich

was chosen as the main test area, but training data was digitised in several cities to improve generalisation.

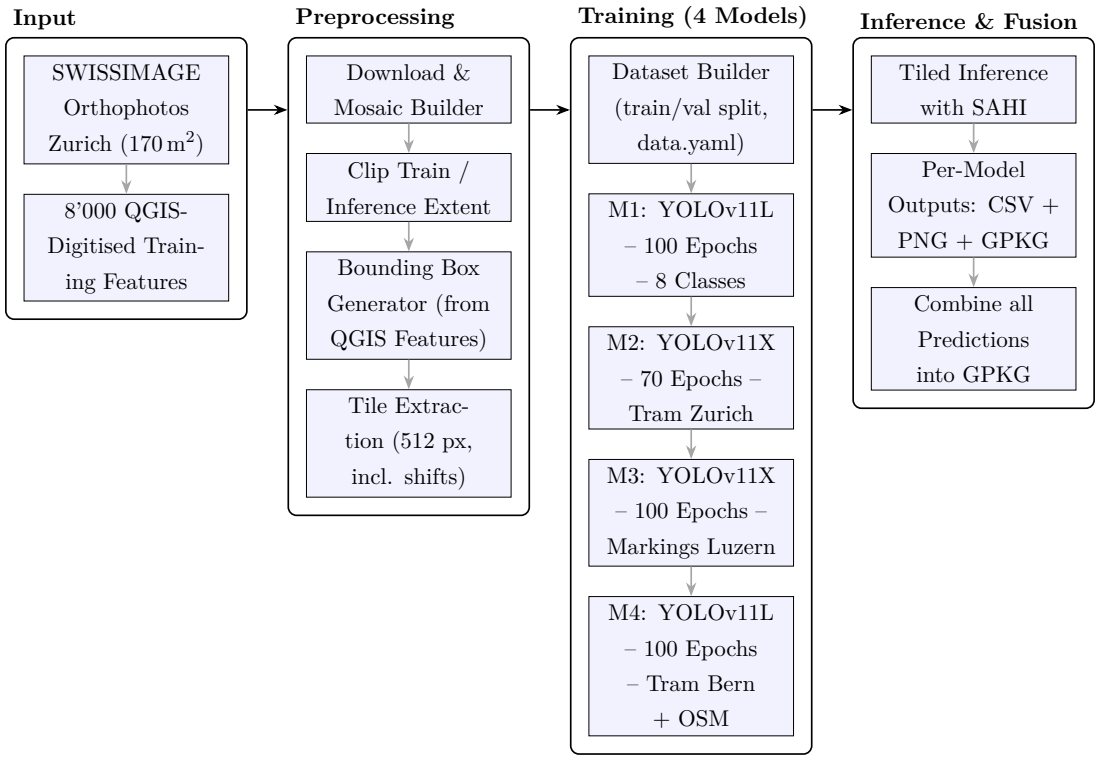
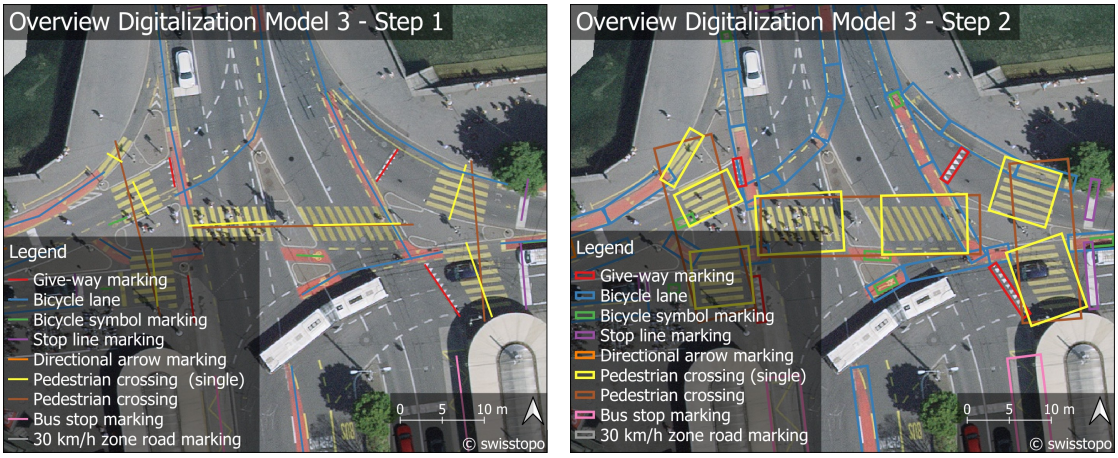


Figure 4.15: Workflow of YOLO object detection on SWISSIMAGE orthophotos (Step 2.5), from QGIS-digitised features and preprocessing to model training, tiled inference with SAHI, and QGIS visualisation.

Model Training

The first stage consisted of creating a high-quality training dataset. In total, around 1,700 vector features were manually digitised in QGIS as part of this study, while an additional 6,300 features were generated semi-automatically based on existing geospatial datasets (see below). This resulted in more than 8,000 annotated objects used for model training. The digitisation process followed a systematic approach: line geometries were drawn for each object type (e.g. tram rails, pedestrian crossings, arrows). A representative class width was then defined, and buffered polygons were created around each line. These polygons were directly exported into YOLO-ready annotations using custom Python scripts. The two-step digitisation workflow is illustrated in Figure 4.16.

Based on these digitised features, the corresponding SWISSIMAGE tiles were identified and downloaded. Instead of using the full swisstopo tile index, a custom Python script developed for this study selected only those tiles overlapping with the digitised geometries. This ensured that only relevant images were downloaded, reducing both storage and processing time. The selected tiles were then merged into larger GeoTIFFs, providing a consistent basis for clipping and tiling in the subsequent stages.



(a) Step 1: manual digitisation of line geometries in QGIS.

(b) Step 2: buffered polygons converted by Python-Script to YOLO format.

Figure 4.16: Two-step digitisation workflow (Lucerne).

The merged GeoTIFFs provided the visual training base for model development. Using these prepared image tiles, a general-purpose model was first created to cover a broad set of urban object classes. All models were trained with the Ultralytics YOLO11 framework on an RTX 4080 GPU (16 GB VRAM). However, some categories proved particularly difficult to detect, such as tram tracks and detailed road markings. For these reasons, three additional specialised models were trained, each focusing on one specific type of object. The four YOLO11 models are summarised below and in Table 4.1.

- **Aerial Model 1 (YOLOv11X, 100 epochs, 8 classes):** This model served as a generalist detector. Approximately 1,000 features were digitised manually in Zurich (not in the City of Zurich), covering the classes *30 km/h zones*, *cars*, *pedestrian crossings*, *arrows*, *school zones*, *tram tracks*, *yield signs*, and *trains*. Several classes, such as tram tracks and pedestrian crossings, proved challenging and were marked for improvement in subsequent models.

- **Aerial Model 2 (YOLOv11X, 70 epochs, tram tracks Zurich/Schlieren/Dietikon):** This specialised model focused on tram infrastructure. Starting from OGP public transport lines (Amt für Raumentwicklung 2025), tram segments were filtered and segmented, with non-visible geometries removed. The training data originated from the tram line between Zurich, Schlieren, and Dietikon, which was selected for its clear visibility and complex network structure.

- **Aerial Model 3 (YOLOv11X, 100 epochs, road markings Luzern):** This model focused on road markings in Luzern and included nine annotated classes: *30 km/h zone marking*, *bus stop*, *pedestrian crossing with island*, *pedestrian crossing without island*, *arrow marking*, *stop marking*, *bicycle marking*, *cycle path marking*, and *yield marking*. About 700 features were manually digitised in QGIS and converted into YOLO annotations.

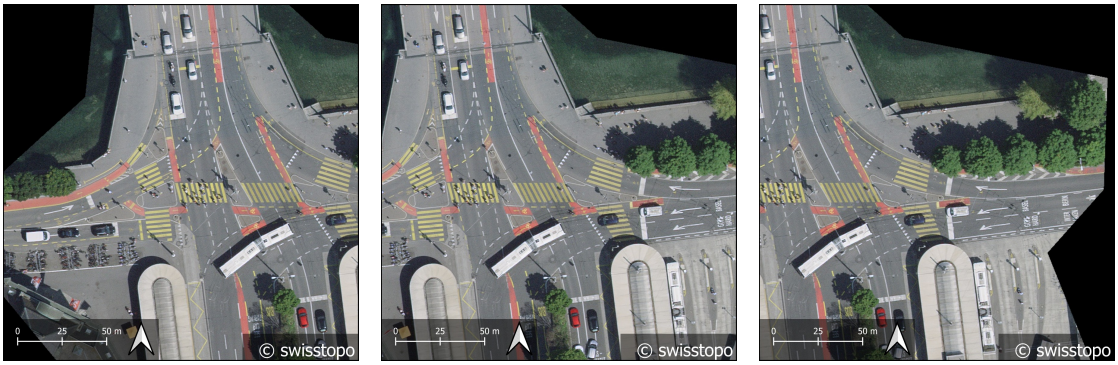
- **Aerial Model 4 (YOLOv11L, 100 epochs, tram tracks Bern):** As tram tracks were the most difficult class to detect, additional training data were collected from Bern, using OSM (OpenStreetMap contributors 2025) as a complementary source. Over 4,000 geometries were segmented and cleaned following the same workflow as for Aerial Model 2.

An overview of all four models and their training configurations is provided in Table 4.1.

Table 4.1: Overview of the four YOLO models trained on SWISSIMAGE orthophotos.

Model	Arch.	Epochs	Classes	Training Data
AM1	YOLO11X	100	8 (mixed)	∼1,000 manually digitised features in Zurich (cars, crossings, arrows, school zones, tram tracks, trains, yield signs, etc.)
AM2	YOLO11X	70	1 (tram tracks)	OGP tram lines between Zurich, Schlieren, and Dietikon; cleaned and segmented
AM3	YOLO11X	100	9 (road markings)	∼700 manually digitised features in Luzern (30 km/h markings, bus stops, pedestrian crossings, arrows, stop markings, bicycle markings, cycle paths, yield markings)
AM4	YOLO11L	100	1 (tram tracks)	OSM tram lines in Bern; segmented and cleaned following the same workflow as AM2

To make the models more robust, data augmentation was applied by creating training tiles of 512 px with different spatial shifts (0.0, 0.3, 0.5, 0.8), see Figure 4.17. This helped to simulate small positional variations and ensured that objects near the tile borders were still well represented in the training data. Rotation-based augmentation did not improve the results and made the later coordinate projection more complex, so it was not used. Shift-based augmentation was therefore chosen as a simpler option. Such geometric transformations are known to improve small-object detection and model generalisation in aerial imagery (Nisa 2024).



(a) Shift = 0.0 (b) Shift = 0.3 (c) Shift = 0.5

Figure 4.17: Illustration of data augmentation using tile shifts. By shifting the clipping window (0.0, 0.3, 0.5), positional noise is simulated and features near tile borders remain represented in the training set.

This shifting strategy simulated positional noise and ensured that features close to tile borders were still well represented in the training set.

4.4 Phase 3: School Route Safety Classification

Phase 3 builds on the outputs from Phases 1 and 2 to estimate a safety score for every edge of the pedestrian network and to make these scores usable for routing. As illustrated in Figure 4.18, this phase integrates multiple analytical steps that can be grouped into three main components: data preparation (Steps 3.1–3.2), safety classification (Steps 3.3–3.4), and routing (Step 3.5).

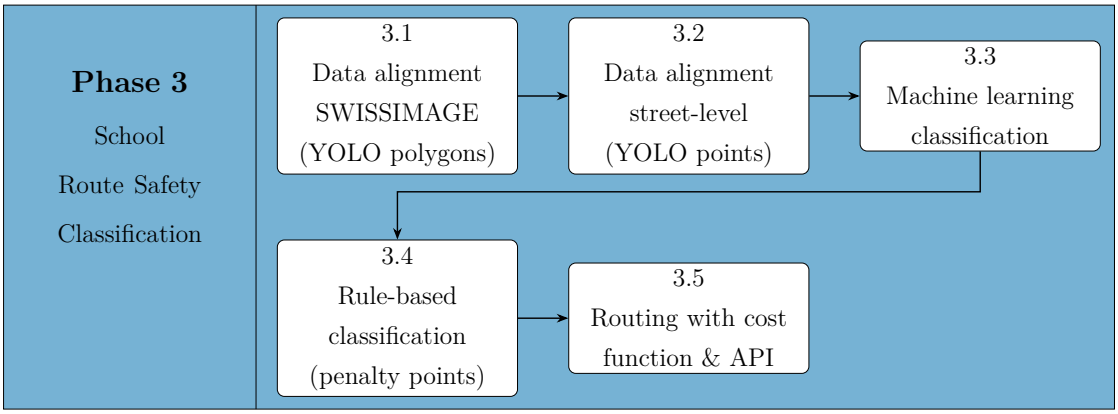


Figure 4.18: Workflow of Phase 3: School Route Safety Classification.

Phase 3 integrates the outputs of both the aerial and street-level computer vision models into the pedestrian network to enable safety classification and routing. Steps 3.1 and 3.2 perform the spatial alignment of these detections with the network graph. This alignment ensures that every network edge contains aggregated safety-related attributes derived from both aerial and ground-level perspectives. The following sections describe the classification approaches and routing implementation in more detail.

Two methods were used to estimate safety. The first is a supervised machine-learning model that predicts how safe each network edge is, using calibration and cross-validation to ensure reliable results (see Section 4.4.3). The second is a rule-based method that assigns penalties and bonuses to specific features, such as missing crossings, tram tracks, or 30 km/h zones (see Section 4.4.4).

The final safety score is converted into routing costs and linked to a routing engine, enabling *safest-path* queries and direct visualisation in QGIS (see Section 4.4.5). To make the results accessible beyond the local environment, the routing engine was integrated into a FastAPI-based web service. This API allows dynamic route computation using the derived safety scores and provides a structured GeoJSON output for further use in GIS applications. A custom QGIS Processing script connects to this API, enabling users to generate, visualise, and compare safe school-route alternatives directly within QGIS (see Section 4.4.5).

The implementation of Phase 3 was carried out entirely in Python. Geospatial processing used `geopandas` (Van den Bossche 2022), `shapely` (Gillies et al. 2025), and `rasterio` (Gillies et al. 2013). Tabular and numerical operations relied on `pandas` (The pandas development team 2020) and `numpy` (Harris et al. 2020). Machine-learning tasks were implemented with `scikit-learn` (Pedregosa et al. 2011), `imbalanced-learn` (Lemaître et al. 2017), and `shap` (Scott M Lundberg et al. 2017) for model interpretation. Parallelisation and progress tracking were supported by `joblib` (Joblib Development Team 2020) and `tqdm` (da Costa-Luis 2019). Further methods for optimisation, statistics, and spatial analysis were provided by `scipy` (Virtanen et al. 2020), and graph operations were handled with `networkx` (Hagberg et al. 2008). API components were built with `fastapi` (Ramírez 2018), `uvicorn` (Encode OSS 2025), and `pydantic` (Colvin et al. 2025). Standard Python modules (`os`, `sys`, `re`, `time`, `gc`, `json`, `warnings`, `unicodedata`, `itertools`, `collections`, `pathlib`) were used for general utilities.

4.4.1 3.1: Data Alignment SWISSIMAGE (YOLO Polygons)

The first step of Phase 3 connected the object detections from aerial imagery with the pedestrian network to create segment-level safety attributes. The workflow of this step is shown in Figure 4.19.

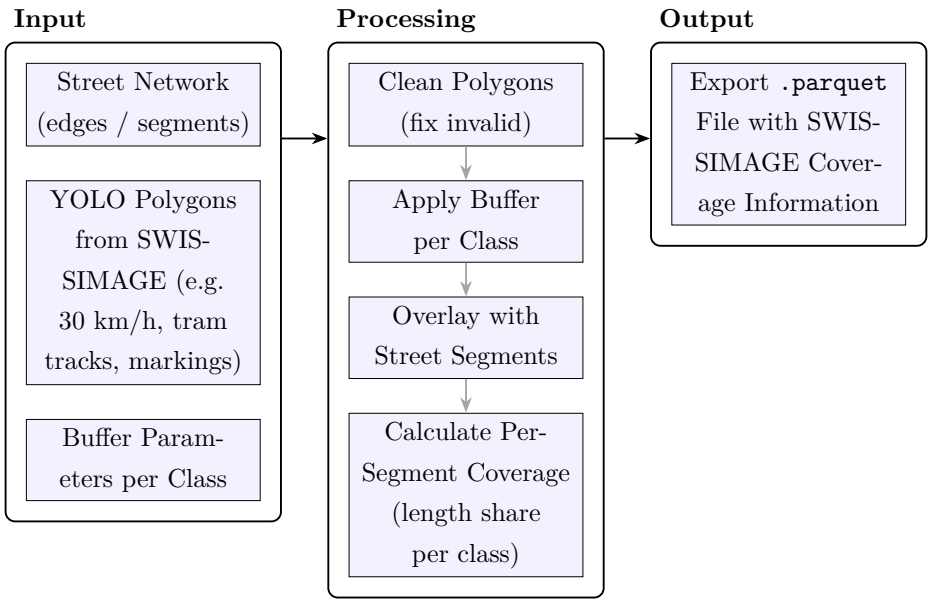


Figure 4.19: Workflow of data alignment (Step 3.1): YOLO polygons are buffered, intersected with street segments, and aggregated into per-class coverage values, exported as a `.parquet` file.

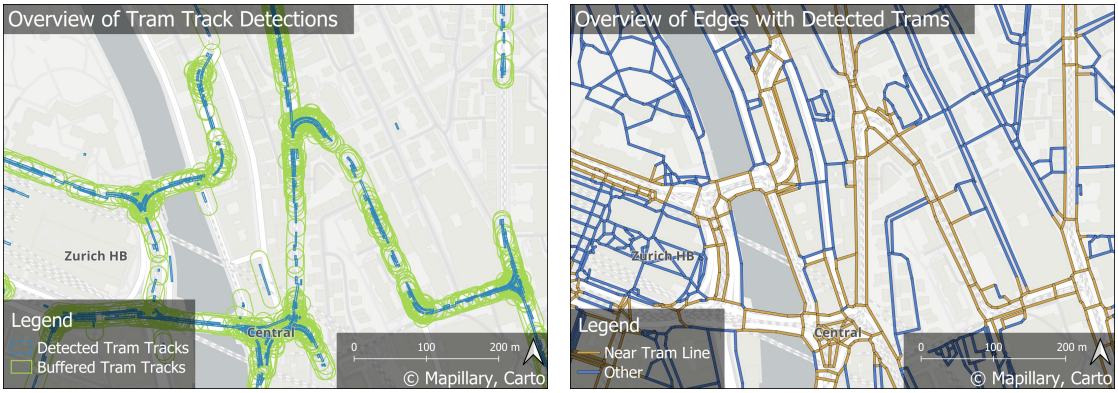
As illustrated in Figure 4.19, the process combined three inputs: (i) the cleaned street network, (ii) YOLO polygon detections from SWISSIMAGE, and (iii) class-specific buffer parameters. Since the raw detections occasionally contained invalid or self-intersecting geometries, all polygons were cleaned first to ensure topological consistency.

Each object class was then buffered to approximate its real-world influence area. Wider buffers were applied to features with broader spatial effects, such as pedestrian crossings, while narrower buffers were sufficient for local pavement markings like arrows. This approach follows common practices in GIS-based feature extraction, where buffers are used to represent the effective zones of influence of objects detected in imagery (Burrough et al. 1998). Table 4.2 summarises the main classes and corresponding buffer distances.

Table 4.2: Buffer parameters applied to YOLO polygon detections from SWISSIMAGE (excerpt).

Class (detection)	**Buffer [m]**	**Notes**
30 km/h zone marking	1.5	Speed-limit zone marking on pavement
Bus stop	4.0	Area around bus stops
Pedestrian crossing (with/without island)	7.0	Wide buffer to capture crossing influence
Stop line	1.0	Stop line before intersections
Yield marking	1.0	Yield / priority markings
Arrow marking	2.0	Directional arrows on pavement
Bicycle lane	2.0	Bicycle lanes (polygon length buffered)
Bicycle symbol	1.0	Bicycle symbol painted on pavement
Tram track	12.0 (+extend 7.0)	Captures corridor of tram rails
Rail track	15.0 (+extend 16.0)	Captures corridor of railway tracks

After buffering, the YOLO polygons were spatially intersected with the street network to link detected features to individual road segments. For each segment, the proportion of its length located within the buffer of each object class was calculated. If this proportion exceeded 20%, a corresponding confidence attribute was assigned to the segment. Segments with a smaller overlap did not receive a confidence attribute. Figure 4.20 illustrates this process using the example of tram tracks: (a) shows the buffered YOLO polygons overlaid on the street network, and (b) displays the resulting attribute values assigned to each segment. Segments with a value of `conf_tram > 0` are labelled as “Near Tram line”.



(a) Buffered YOLO detections overlaid on the street network.

(b) Segments with `conf_tram > 0` (*Near Tram line*).

Figure 4.20: Example of data alignment and attribute assignment.

The enriched network containing these per-segment coverage shares was exported as a `.parquet` file. This dataset forms the common basis for the subsequent safety classification approaches, namely the supervised machine-learning model (Section 4.4.3) and the rule-based scoring scheme (Section 4.4.4).

4.4.2 3.2: Data Alignment Street-Level YOLO Points

Street-level detections complemented the aerial data by adding ground-level information on traffic and infrastructure features. The purpose of this step was to convert the raw point detections into spatial indicators that could be linked to the pedestrian network. The workflow of this process is shown in Figure 4.21.

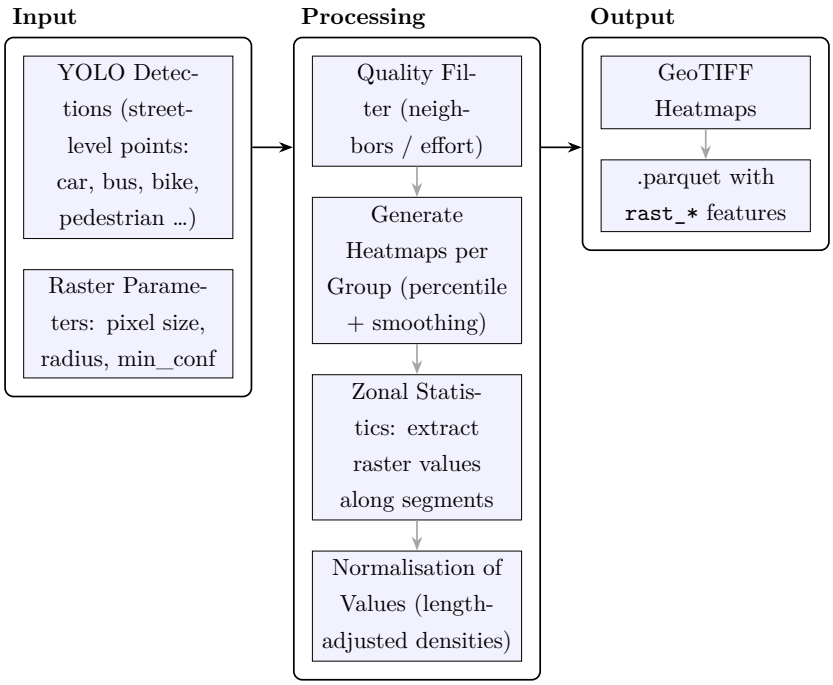


Figure 4.21: Workflow of data alignment (Step 3.2).

Since the raw detections contained hundreds of fine-grained labels from the Mapillary dataset, they were first aggregated into broader, safety-relevant groups. Table 4.3 summarises the main categories used in this process.

Table 4.3: Grouping scheme used for street-level detections.

Group name	**Included raw labels**
Associated with Bikes	Bike symbols and markings, bicycles, cycle lanes, cyclists, traffic lights for cyclists
Arrows	All painted directional arrows (left, right, straight, combined)
Bridges	Bridge structures over roads or railways
CCTV Cameras	Surveillance and traffic cameras
Construction / Parking	Parking signs, parking surfaces, aisles, and meters
Construction / Road Elements	Curbs, fences, barriers, walls, junction boxes, overpasses, other fixed roadside structures
Crosswalks	Zebra crossings and other pedestrian crosswalks
Dashed Markings	Broken or dashed road lines
Humans / Pedestrians	People (individuals or groups) on or near the road
Other Markings	Give-way rows, solid or zigzag lines, hatched/chevron markings, and miscellaneous symbols
Street Furniture / Nature	Vegetation, benches, bike racks, small roadside furniture
Other Objects	Traffic cones, buildings, manholes, hydrants, ground text
Pedestrian Areas	Sidewalks, pedestrian zones, and plazas
Road Surfaces	Roads, shoulders, service lanes, sidewalks, curb cuts
Signs (General)	Information signs, backplates, miscellaneous traffic signs, and traffic lights (other)
Stop Lines	Painted stop lines
Street Lights	Lampposts and other public lighting
Traffic Islands	Raised or painted traffic islands
Traffic Lights	Standard traffic lights for vehicles and pedestrians
Traffic Signs	Traffic direction, regulatory, and temporary signs, including frames and supports
Tram / Rail	Tram and rail tracks embedded in the road
Tunnels	Tunnel entrances and covered passages

Table 4.3: Grouping scheme used for street-level detections (continued)

Group name	**Included raw labels**
Vehicles (Motorised)	Cars, buses, trucks, motorcycles, trailers, slow vehicles, trams

After grouping the detections, several processing steps were applied to clean and combine the results. The aim was to produce reliable spatial representations that indicate where specific features are likely to occur, based on both the number of detections and their model confidence values.

For each feature group, specific parameters were defined, including the *kernel radius*, *minimum confidence*, and *minimum local effort* (Table 4.4). Detections below the confidence threshold or located in images without sufficient spatial coverage (*minimum local effort*) were removed. The *local effort* criterion required a minimum number of neighbouring images within a 15,m radius, which was determined empirically as the most stable balance between noise reduction and spatial detail.

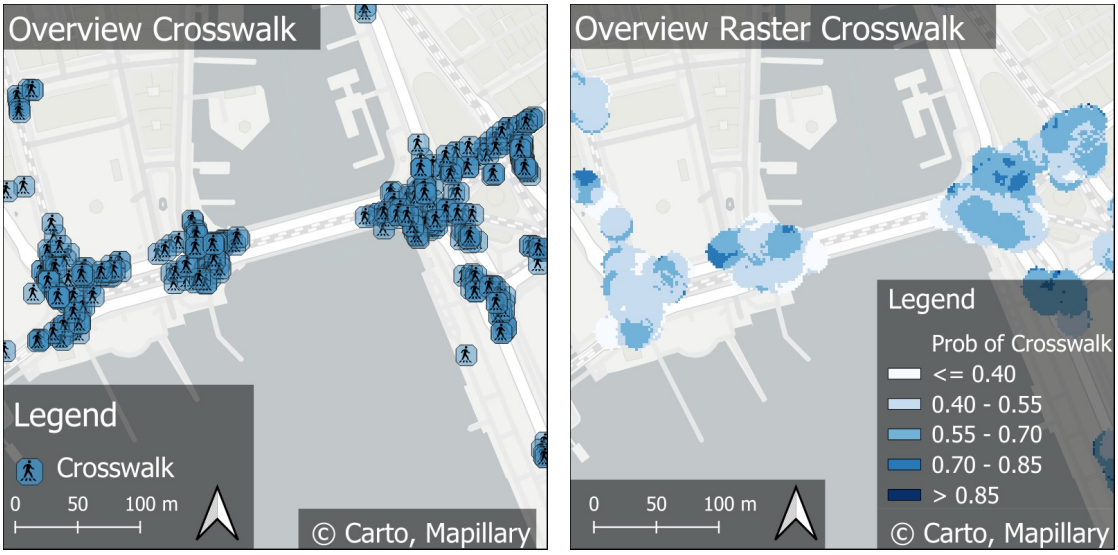
Table 4.4: Selected group-specific rasterisation parameters used for smoothing and filtering.

Group	**Kernel radius [m]**	**Min. confidence**	**Min. local effort**	**Description**
Arrow	2.0	0.2	1	Directional road arrows
Bicycle Infrastructure	6.0	0.3	2	Bike lanes, racks, and cycle markings
Bridge / Tunnel	12.0	0.2	2	Over- and underpasses
Crosswalk	3.0	0.2	2	Pedestrian crossings and zebra markings
Nature / Vegetation	12.0	0.2	2	Trees, bushes, and natural areas
Stop Line	1.0	0.2	1	Very localised markings near intersections
Traffic Lights	6.0	0.4	2	Signals for vehicles and pedestrians
Vehicles (motorized)	8.0	0.3	1	Motor vehicles and on-road traffic

Each valid detection was then assigned a circular buffer corresponding to its group-specific *kernel radius*, representing the spatial influence of that object. All buffers were

rasterised on a 2,m grid (EPSG:2056). For each grid cell, the 95th percentile of all overlapping detection confidences was calculated. Using the percentile instead of a mean or maximum value emphasised areas with consistently strong detections and reduced the effect of isolated or uncertain ones. The resulting rasters thus represent continuous confidence surfaces rather than discrete detection points.

Figure 4.22 illustrates this process for the *crosswalk* group. Panel (a) shows the original Mapillary detections, while panel (b) shows the resulting smoothed confidence surface. High-confidence areas form coherent zones reflecting the true spatial extent of crossings, whereas isolated, low-confidence points are filtered out during the smoothing process.



(a) Crosswalk detections from Mapillary (points)

(b) Confidence surface of crosswalk detections (2 m, smoothed)

Figure 4.22: Example of rasterisation and confidence smoothing for pedestrian crossings.

Conceptually, this conversion of discrete detections into continuous surfaces follows the principle of kernel-based spatial smoothing (Davies et al. 2017). As Laube et al. (2008) describe, spatial influence can be understood as a fuzzy relationship that decreases with distance. This concept was adapted here by treating detection confidences as spatially continuous values that decline radially from each object, resulting in *spatial confidence surfaces* that capture both detection reliability and spatial proximity.

In the final step, the confidence rasters were linked to the pedestrian network. For each network edge, the mean confidence within a 12,m buffer was calculated. This buffer distance represents the typical spatial range visible in street-level imagery and captures nearby features—such as crossings, traffic lights, or vegetation—that affect walking conditions. The resulting average confidence values were stored as additional attributes for each feature group, providing context-aware measures of the surrounding environment. This spatial attribution follows established GIS exposure-modelling principles, where environmental information is aggregated within local neighbourhoods or buffer zones (Ali et al. 2002). Similar to the spatial filtering approach described by Ali et al. (2002),

this method integrates nearby influences to obtain smoothed, context-aware measures around each network edge.

4.4.3 3.3 Machine Learning Classification

After aligning and aggregating the aerial and street-level detections to the pedestrian network, the resulting attributes provided a detailed spatial description of safety-relevant features for each edge. In this step, these network-based data were used to estimate safety levels through a supervised machine learning framework.

The goal was to predict how safe each street segment is, based on existing empirical assessments. The task was set up as an *ordinal classification* problem: given segment-level features, the model estimated the probability that a segment belonged to one of four difficulty levels defined by the Zurich City Police (Stadtpolizei Zürich 2025).

The input data combined results from Steps 3.1 and 3.2, including coverage shares from *SWISSIMAGE* polygons, raster-based indicators derived from street-level imagery, and network attributes such as edge length. Together, these features described both the visual and structural characteristics of the pedestrian environment. As ground truth, police assessments of crossing difficulty were used (Stadtpolizei Zürich 2025). These labels were defined on an ordinal scale with four categories: “suitable”, “increased requirement”, “demanding”, and “not recommended”.

Tree-based ensemble models were chosen for the classification. They combine many individual decision trees into one robust predictive model. Each tree is trained on a random subset of the data and features, and the final prediction is obtained through majority voting (Breiman 2001). This method reduces overfitting, handles noisy or incomplete data well, and performs reliably on mixed tabular datasets without assuming specific data distributions.

To make the model interpretable, SHAP (SHapley Additive exPlanations) values were calculated (Scott M. Lundberg et al. 2019). SHAP explains how much each feature contributes to a prediction by comparing model outputs with and without that feature.

Because the dataset was strongly imbalanced, with only about 30 samples in the smallest class, synthetic oversampling was applied using the SMOTE algorithm (Chawla et al. 2002). SMOTE creates artificial examples for minority classes and improves model performance in unbalanced datasets, especially when used with tree-based methods such as Random Forests (Desprez et al. 2022). Balanced accuracy was used as the main evaluation metric because it treats all classes equally and provides a fairer measure of model performance on imbalanced data (Brodersen et al. 2010).

Finally, the ordinal structure of the labels was preserved in the prediction stage. Expected values over the ordered classes were used to produce continuous safety scores, and threshold tuning was applied to improve the separation between difficulty levels. The complete workflow of data preparation, model training, and prediction is shown in Figure 4.23.

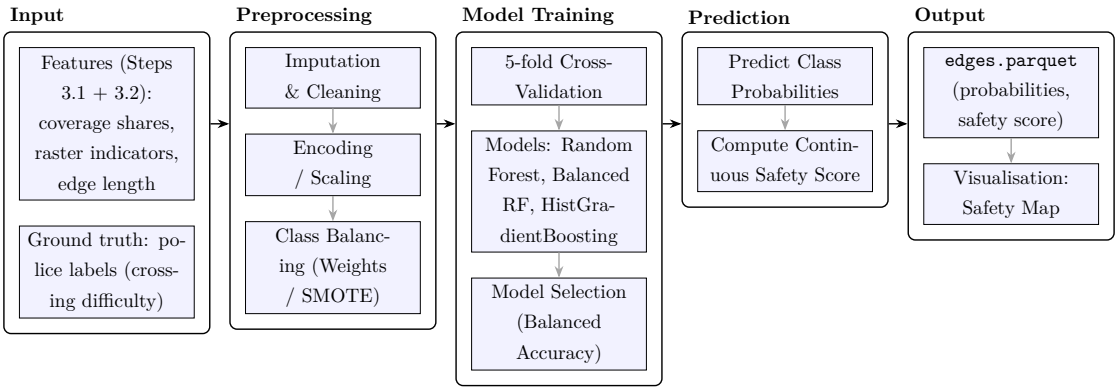


Figure 4.23: Workflow of the machine learning classification (Step 3.3).

The cleaned dataset contained 102,590 network segments, of which 2,310 had ground truth labels for training. The class distribution was strongly imbalanced, with 1,811 segments rated as "suitable", 355 as "increased requirement", 114 as "demanding", and only 30 as "not recommended".

Four ensemble models were compared: a Random Forest with random oversampling, a Random Forest with SMOTE (Chawla et al. 2002; Breiman 2001), a Balanced Random Forest (Breiman 2001), and a Histogram-based Gradient Boosting classifier (Guryanov 2019). All models were evaluated using 5-fold stratified cross-validation (Gorriz et al. 2024). This approach provides a balanced compromise between bias and variance, ensuring that every sample is used for both training and validation while maintaining the original class proportions. According to Gorriz et al. (2024), K-fold Cross-Validation remains one of the most robust techniques for model validation in heterogeneous and limited datasets, offering reliable error estimates without assuming specific data distributions. A 5-fold scheme was selected to stabilise performance estimates and to avoid the high variance associated with leave-one-out validation when sample sizes are small Gorriz et al. (2024).

The Random Forest with SMOTE achieved the highest balanced accuracy (0.478) and was therefore selected as the final classifier. The corresponding macro-F1 was 0.484 and the weighted-F1 0.789, confirming that the model captured all four classes with acceptable discrimination.

A separate hold-out evaluation using 20% of the data was performed to verify model stability. This external validation step provides an independent estimate of the generalisation error using data unseen during model training or cross-validation (Kuhn et al. 2013). According to Kuhn et al. (2013), such hold-out testing offers an unbiased measure of predictive performance and serves as a final confirmation of model reliability after cross-validation. The hold-out results closely matched the cross-validation findings, with a balanced accuracy of 0.447 and a macro-F1 of 0.433, indicating consistent performance and no evidence of overfitting.

After per-class threshold tuning, the balanced accuracy increased to 0.486 and the macro-F1 to 0.450. Threshold tuning adjusted the class probabilities to improve the

recognition of rare categories while preserving the ordinal label structure. This adjustment led to higher recall for minority classes such as “demanding” and “not recommended”, while the overall accuracy remained around 0.8 due to the dominance of the majority class “suitable”. These results demonstrate that the model identified difficult crossings more reliably without degrading its performance on safe segments.

Finally, the selected Random Forest with SMOTE was retrained on all available labelled data and applied to the complete set of 102,590 network segments. For each segment, class probabilities (`prob_*`) and a continuous safety score were computed.

Figure 4.24: Illustrative map of ML-based safety classification: police labels (four ordinal categories) were used to train the model, which predicts per-segment probabilities and continuous safety scores.

To obtain interpretable and continuous safety values, the predicted class probabilities were transformed into a numerical safety score using a weighted expected-value formulation:

$$\text{safety\_score} = \sum_i w_i \cdot P_i,$$

where w_i denotes the class weight and P_i the predicted probability for class i. For example, if a segment was predicted with $P(\textit{suitable}) = 0.7$, $P(\textit{increased requirement}) = 0.2$, $P(\textit{demanding}) = 0.08$, and $P(\textit{not recommended}) = 0.02$, the resulting safety score would be

$$\text{safety\_score} = (0.7 \times 100) + (0.2 \times 20) + (0.08 \times 10) + (0.02 \times 1) = 76.2.$$

This value corresponds to a segment that is predicted to be generally safe, while a small proportion of uncertainty remains due to minor probabilities of more difficult classes.

This expected-value transformation is a well-established approach in probabilistic modelling, as it converts discrete class probabilities into a single continuous score while preserving the ordinal structure of the labels. A comparable concept has been applied in the medical domain, where ordinal diagnostic labels (e.g. disease severity levels) are converted into continuous outcome values to capture subtle variations in prediction confidence (Hoebel et al. 2023). Although developed for clinical scoring, the underlying mathematical principle is identical to the approach used in this thesis: both estimate the expected value of an ordered categorical variable from predicted class probabilities, thereby providing a more nuanced and interpretable representation of model output.

In this thesis, the following ordinal weights were used to represent decreasing levels of pedestrian safety: "suitable" = 100, "increased requirement" = 20, "demanding" = 10, and "not recommended" = 1. This weighting scheme was chosen to ensure interpretability while maintaining a clear ordinal separation between categories. By compressing the lower end of the scale, the influence of rare unsafe cases is reduced, thereby preventing overrepresentation of highly uncertain predictions in areas with limited training data.

The use of a continuous safety index serves two main purposes. First, it captures subtle differences between segments by incorporating the full probability distribution rather than relying on a single discrete prediction. Second, it enables direct comparison with the rule-based classification described in Section 4.4.4, as both methods produce results on the same 0–100 scale.

After the safety classification, the data was saved in a .parquet file.

4.4.4 3.4: Rule-Based Safety Classification

In parallel to the machine-learning models, a transparent and interpretable rule-based safety classification was developed. The goal of this method was to turn basic traffic-safety principles into a simple numerical score that can be compared with the machine-learning results. Unlike the data-driven models, it was created manually through step-by-step testing and adjustment, so that the system reflects both theoretical knowledge and practical experience. The idea was to make the logic as clear as possible: every feature either increases or decreases the overall risk in a way that is easy to understand. An overview of this process and the main calculation steps of the rule-based model is illustrated in Figure 4.25.

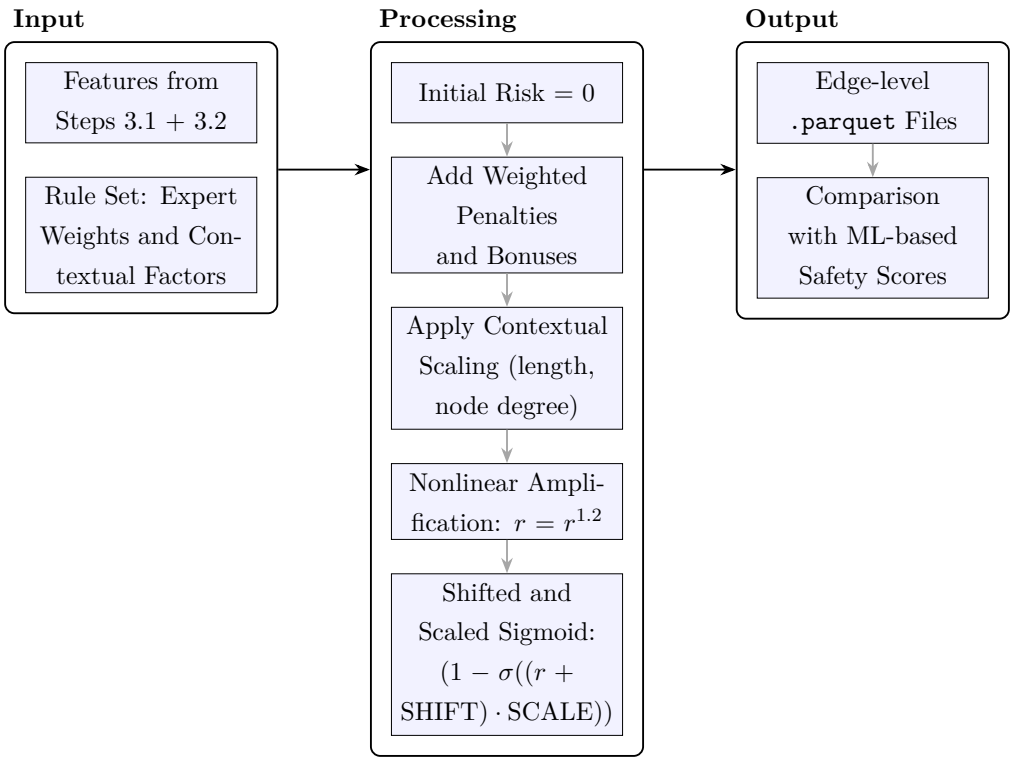


Figure 4.25: Workflow of the rule-based safety classification (Step 3.4): expert-defined weights and contextual scaling factors are combined into a continuous, interpretable safety score for each network segment.

Each street segment starts with a neutral risk value of 0. Feature-based penalties and bonuses are then added using fixed weights. Following the approach described by Leur et al. (2002), the weights were chosen based on the traffic-safety literature and insights from Chapter "Conceptualising School-Route Safety", and were adjusted through repeated experimentation and visual inspection of the results. In practice, higher penalties were assigned to features that clearly increase pedestrian risk (such as tram lines, bus stops, or missing crossings), while bonuses were used for protective elements (such as 30 km/h zones, marked crossings, or refuge islands). The full set of penalty and bonus weights is listed in Tables 4.5 and 4.6.

After all features are added, the accumulated risk value r is scaled by segment length and intersection complexity (node degree), so that long or complex segments with many connections receive slightly higher values. To make the differences between very safe and very unsafe segments more visible, the risk is then slightly amplified by a non-linear transformation $r = r^{1.2}$. This amplification increases the contrast between highly safe and highly unsafe segments, reflecting the empirical observation that perceived danger and crash risk often grow faster than linearly with environmental complexity (Leur et al. 2002).

Before applying the final mapping, the distribution of the accumulated raw risk values (r_{raw}) was inspected. Most segments exhibited low risk values, while a small number reached noticeably higher levels. To obtain a bounded and continuous safety scale, a

shifted and scaled sigmoid transformation was applied to the accumulated risk values.

Let $\sigma : \mathbb{R} \to (0, 1)$ denote the logistic sigmoid function as illustrated in Figure 4.26. , defined as

$$\sigma(x) = \frac{1}{1 + e^{-x}}.$$

The final safety score s for each segment is then computed as

$$s(r) = \big(1 - \sigma((r + \text{SHIFT}) \cdot \text{SCALE})\big) \cdot 100,$$

with SHIFT $= -5$ and SCALE $= 0.5$. Here, $r \in \mathbb{R}_{\geq 0}$ represents the accumulated risk after applying all feature-based penalties, contextual scaling, and non-linear amplification. The parameters *SHIFT* and *SCALE* control the horizontal offset and steepness of the function, respectively.

This transformation follows the classical Verhulst logistic sigmoid function (Kyurkchiev et al. 2015), which provides a smooth and bounded mapping from unbounded input values r to a fixed output range $s \in [0, 100]$. Shifted logistic functions are widely used to approximate threshold-like relationships in a continuous way. They offer numerical stability and interpretability while reducing the influence of extreme values. Compared to a linear normalization, this transformation provides a smoother and more realistic transition between safe and unsafe conditions, while preserving high sensitivity in the range most relevant for children and pedestrians. Such bounded and non-linear mappings are widely used in composite safety indices and risk normalization frameworks (Iranitalab et al. 2017; Silva et al. 2020), as they combine interpretability with numerical stability and ensure that the resulting scores remain within a consistent and interpretable 0–100 scale.

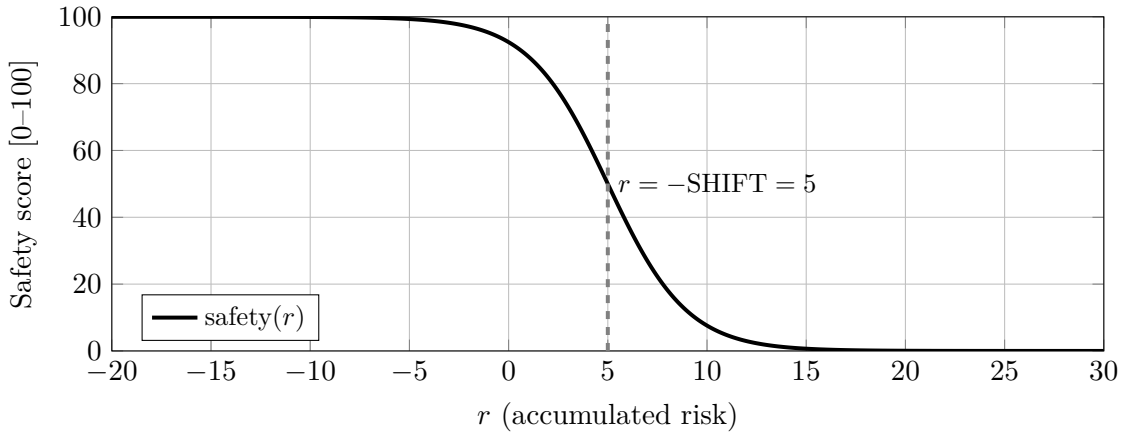


Figure 4.26: Logistic mapping of accumulated risk r to a bounded safety score. The black curve follows the shifted Verhulst sigmoid function with parameters SHIFT $= -5$ and SCALE $= 0.5$, where higher r values correspond to lower safety.

The parameters and weights were tuned by trial and error to produce realistic values. For example, a short residential dead-end street with a 30 km/h zone and a marked crossing typically receives a score around 85, while a wide intersection with a tram line

and no pedestrian markings may fall below 30. These results were visually inspected on the map and adjusted until the scoring pattern reflected real-world conditions as closely as possible.

Table 4.5: Rule-based classification: penalty weights (positive values increase risk). Raster-detected features refer to objects automatically extracted from Mapillary imagery.

Condition / Feature	**Weight**	**Notes**
Crossing without marking	+3.0	Unmarked pedestrian crossing; high risk for children.
Crossing without pedestrian priority	+1.2	No pedestrian right-of-way.
Crossing with tram	+3.0	Crossing intersects with a tram line.
Crossing with tram (no pedestrian protection)	+6.0	Tram crossing without pedestrian protection (most critical).
Crossing with tram (island or marked)	+1.2	Risk mitigated by refuge island or pavement markings.
Car traffic	+0.4	General exposure to motorized traffic.
Bus stop	+0.3	Increased risk due to bus maneuvers and passenger activity.
Node complexity (in/out degree)	+0.008	Higher intersection complexity increases potential exposure.
Tram (gated)	+2.0	Tram line present at pedestrian crossing (gated condition).
Tram rail (gated)	+1.2	Tram rail present at gated pedestrian crossing.
Bicycle lane	+0.5	Adjacent bicycle lane; potential conflicts for small children.
Bicycle marking	+0.3	Painted bicycle markings indicate mixed traffic conditions.
Raster: tram rail	+0.4	Mapillary-detected tram rail.
Raster: motorized vehicles	+0.5	Mapillary-detected motorized vehicles on the road.
Raster: dashed marking	+0.3	Mapillary-detected dashed road marking.

Table 4.6: Rule-based classification: bonus weights (negative values reduce risk). Raster-detected features refer to objects automatically extracted from Mapillary imagery.

Condition / Feature	**Weight**	**Notes**
30 km/h zone	−0.5	Speed-limit area (slower traffic, safer conditions).
30 km/h marking	−0.2	Visible 30 km/h road marking.
Marked pedestrian crossing	−0.5	Presence of a pedestrian crossing.
Pedestrian crossing with island	−0.8	Crossing with refuge island.
Stop marking	−0.3	Stop marking increases driver awareness.
Raster: crosswalk	−0.05	Mapillary-detected pedestrian crossing (weak evidence).
Raster: traffic lights	−0.05	Mapillary-detected traffic lights.
Raster: refuge island	−0.025	Mapillary-detected refuge island.
Raster: stop line	−0.025	Mapillary-detected stop line.
Raster: street lighting	−0.0125	Mapillary-detected street lighting.

Overall, this rule-based model was designed to be simple, transparent, and easy to interpret. The final scores were saved in edge-level `.parquet` files for direct comparison with the machine-learning predictions.

4.4.5 3.5: Routing with Cost Function

The segment-level safety scores obtained in Steps 4.4.3 and 4.4.4 were integrated into a routing system. An overview of this process is shown in Figure 4.27. The goal of this component was to transform abstract safety indicators into a practical decision-support tool that could generate safer pedestrian routes between any two points. To achieve this, a method was implemented that directly incorporated safety values into the cost function used for path finding.

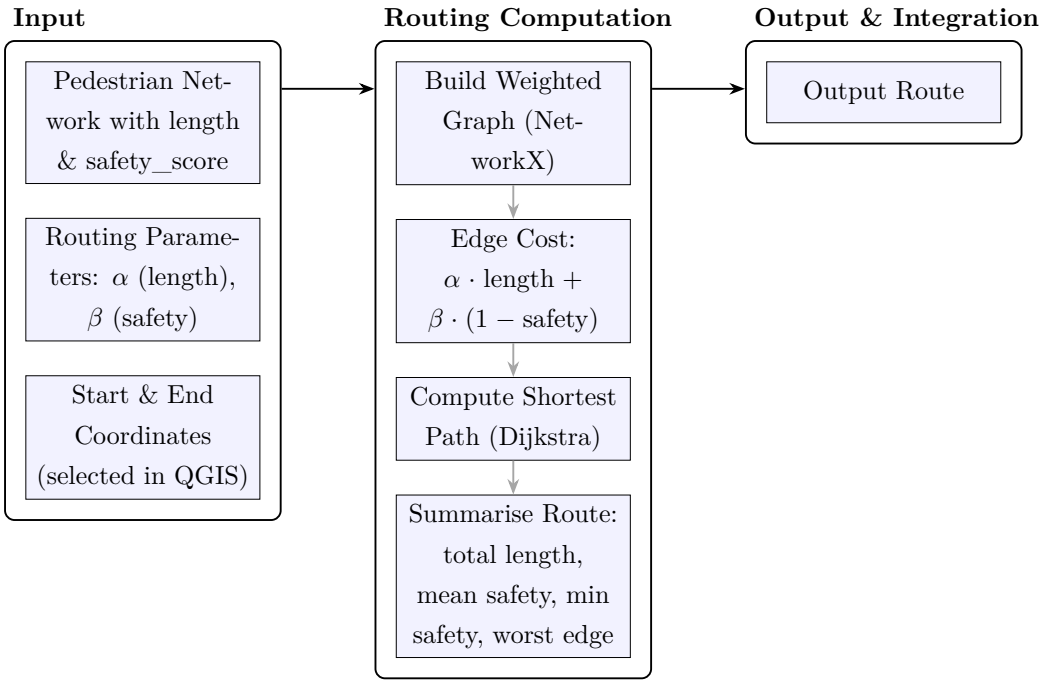


Figure 4.27: Routing workflow showing how length and safety are combined to compute the optimal path within the API.

The routing graph was based on the pedestrian network, where each edge was annotated with both its length and safety score. A combined cost function was defined as

$$\text{cost} = \alpha \cdot \text{length} + \beta \cdot (1 - \text{safety}),$$

with the parameters α and β controlling the trade-off between distance and safety. This general formulation follows the multi-criteria route evaluation principle proposed by Völkel et al. (2008), who developed a detailed weighting model for pedestrian route choice that integrates both objective and subjective factors such as surface quality, lighting, traffic volume, and safety perception. In contrast to Völkel et al. (2008) detailed model based on survey data and qualitative criteria, this thesis applies a simplified, data-driven approach. Here, the safety term is derived from automatically detected environmental features (see Steps 4.4.3–4.4.4) and scaled by edge length to ensure that longer unsafe segments contribute proportionally more to the total cost. This simplification enables reproducible computation of large-scale school route networks while remaining conceptually consistent with the foundations of Völkel et al. (2008).

Each route was summarised with key indicators such as total length, mean safety, minimum safety, and the attributes of the "worst" edge along the path. This ensured that route suggestions were not only technically optimal but also interpretable for planners and stakeholders.

To make the routing functionality usable outside the analytical environment, a simple `FastAPI` web service was created (Ramírez 2018). It loads the pedestrian network with the computed `safety_score` values and provides a single endpoint (`/route`) that

calculates routes between any two points. Users can specify the weighting parameters (α, β) to decide how strongly distance or safety should influence the result. Internally, the service builds a `NetworkX` graph (Hagberg et al. 2008), applies the cost function, and returns the route as a GeoJSON file containing the geometry and summary values such as total length and average safety. The API can be started locally or via Docker and includes an interactive `Swagger UI` for quick testing.

For practical use, the API was connected to QGIS through a custom Processing Script (`QGIS_Script.py`). This allows users to calculate safe routes directly in the QGIS interface: start and end points are selected on the map, the parameters (α, β) are defined, and the script automatically requests the result from the API. The returned GeoJSON route is then added as a new map layer automatically. In this way, the routing system becomes transparent, easy to reproduce, and directly applicable in everyday planning work without any programming.

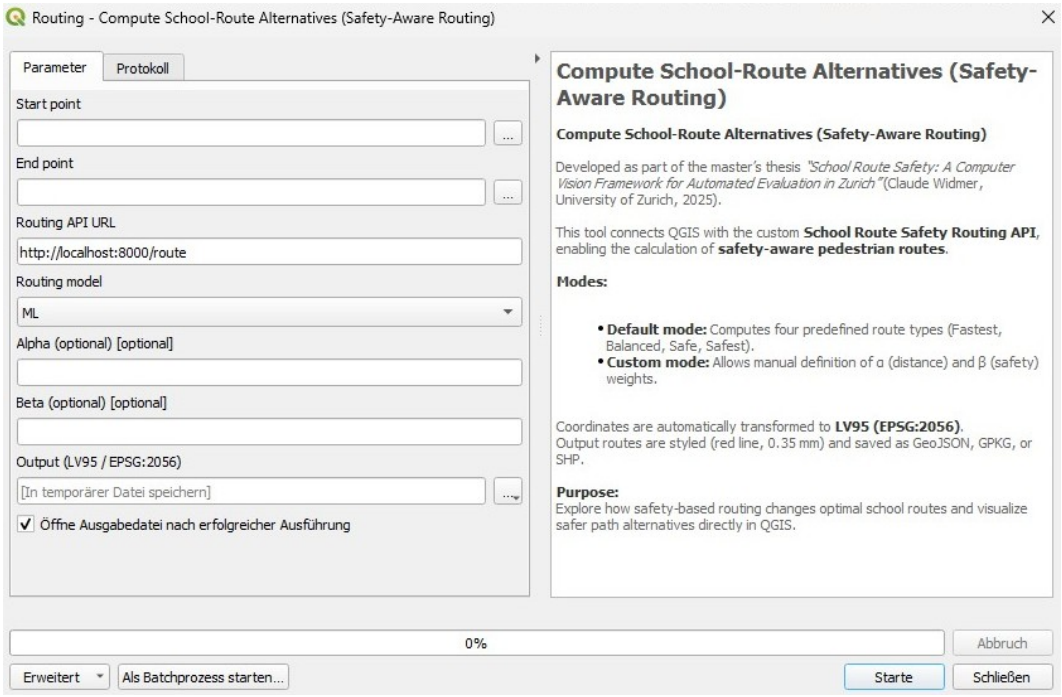


Figure 4.28: QGIS interface of the routing script: users can define start and end points, adjust route weights (α, β), and display the resulting safe route directly on the map.

The combined API–QGIS setup links the analytical backend with a visual and intuitive interface.

4.5 Phase 4: Visualization

The final phase focused on the visualization of all results in both static and interactive formats. Visual analyses were implemented in QGIS (QGIS Development Team 2025) for cartographic presentation and interpretation of spatial patterns, while complementary

analytical plots and map overlays were generated with **Python**. These visualizations supported the interpretation of spatial safety variations and enabled the empirical answering of the research questions defined in Section Research Objectives and Questions. An overview of the visualization workflow is shown in Figure 4.29.

Figure 4.29: Workflow of Phase 4: Visualization.

5 Results

This chapter presents the empirical findings of the study and directly addresses the four research questions (RQ1–RQ4) introduced in Chapter 1. The structure of this chapter follows the methodological workflow outlined in Chapter 4. Where appropriate, key intermediate steps are also reported to provide transparency on how the final results were obtained. Together, these findings form the empirical foundation for evaluating the overall research aim:

> *To design and evaluate an automated, computer-vision–based framework to systematically assess the safety of school routes in the City of Zurich.*

The interpretation and broader implications of these results are discussed in Chapter 6.

5.1 Pedestrian Network Preparation Results

The first part of the analysis focuses on preparing the pedestrian network, which serves as the spatial basis for all subsequent steps. This section presents the main results of the network cleaning and integration process described in Section 4.2, including the extent of the final network, improvements in connectivity, and the computational performance of the workflow.

5.1.1 Network Extent and Topology

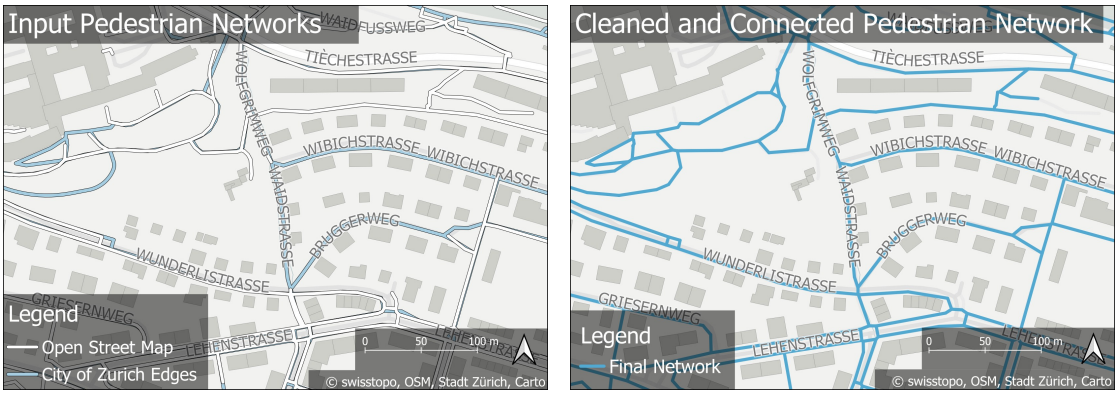
Table 5.1 summarises the overall extent and structural characteristics of the three pedestrian networks. The filtered OSM network covered a total length of approximately 1940 km, while the municipal pedestrian dataset of the City of Zurich reached about 1860 km. After merging and topological cleaning, the combined network resulted in a total length of 2353 km, representing an increase of roughly 20–25 % in overall coverage. The mean node degree of the final network was 2.45, indicating a stable and consistent network structure.

Table 5.1: Overall network extent and structure before and after merging.

Dataset	**Total length [km]**	**Mean node degree**
OSM Filtered	1939.7	1.43
City of Zurich	1864.4	2.92
Final Combined	2353.4	2.45

Figures 5.1a and 5.1b illustrate the spatial extent of the source networks and the resulting merged dataset. The final network integrates both the dense local paths from

OSM and the geometrically precise main corridors from the municipal dataset, providing full coverage of the study area. For example, smaller residential connections such as the *Bruggerweg*, which were missing in the OSM dataset, are now included in the final pedestrian network.



(a) Input pedestrian networks (OSM and City of Zurich).

(b) Final merged and cleaned pedestrian network.

Figure 5.1: Comparison of input and processed pedestrian networks.

5.2 Detected Features Performance

Building on the prepared network, the next part of the analysis evaluates the detection performance of the developed computer-vision models. Two model types are assessed: (i) the *street-level model*, trained on Mapillary images, and (ii) the *aerial model*, trained on SWISSIMAGE orthophotos.

5.2.1 Street-level Models

The following subsections analyse the performance of the street-level detection models trained on Mapillary imagery. A total of four YOLO11 variants were trained on subsets of the Mapillary Vistas dataset, differing in architecture and training configuration (see 4.3.2). Their performance is evaluated in terms of both bounding-box detection and instance segmentation, providing insights into overall accuracy and robustness across classes.

Training and Overall Metrics

Tables 5.2 and 5.3 provide a comparison of all four model variants. Table 5.2 reports precision, recall, and mean average precision at IoU thresholds of 0.50 (mAP_{50}) and 0.50–0.95 ($mAP_{50\text{–}95}$), together with the corresponding validation losses. IoU (Intersection over Union) quantifies the overlap between predicted and true object regions.

Table 5.2: Bounding box metrics for YOLO11 Mapillary models (last epoch).

Model	Precision	Recall	mAP_{50}	$mAP_{50\text{–}95}$	$\frac{\text{val}}{\text{box\_loss}}$
Model 1	0.49	0.13	0.13	0.08	1.55
Model 2	0.42	0.22	0.21	0.14	1.27
Model 3	0.49	0.27	0.28	0.19	1.18
Model 4	0.50	0.28	0.29	0.19	1.22

Table 5.3: Segmentation metrics for YOLO11 Mapillary models (last epoch).

Model	Precision	Recall	mAP_{50}	$mAP_{50\text{–}95}$	$\frac{\text{val}}{\text{seg\_loss}}$
Model 1	0.47	0.10	0.10	0.05	3.39
Model 2	0.40	0.17	0.16	0.08	2.78
Model 3	0.45	0.22	0.22	0.11	2.58
Model 4	0.48	0.23	0.23	0.12	2.62

Street-Level–Models 3 and 4 achieved the best overall balance between precision and recall, with the highest mean average precision scores (mAP_{50} up to 0.29 and $mAP_{50\text{–}95}$ up to 0.19) and the lowest validation losses.

Per-class Detection Results

In addition to the overall model metrics, per-class results were computed. Table 5.4 reports the precision, recall, and AP_{50} values for a selection of safety-relevant classes in Street-Level–Model 3. This includes features such as crosswalks, sidewalks, traffic lights, traffic signs, and vehicles, which are of particular interest for street-level safety analysis.

For bounding box predictions, the table lists per-class precision, recall, and AP_{50} values, while for segmentation results, the corresponding precision, recall, and F1-scores are provided. This excerpt illustrates representative categories, while the complete set of per-class results for all four models is provided in the appendix.

Table 5.4: Per-class results for selected safety-relevant classes (Street-Level–Model 3, Bounding Box and Segmentation).

Class	Bounding Box			Segmentation		
	Precision	Recall	AP_{50}	Precision	Recall	F1
construction–flat–bike-lane	0.27	0.29	0.14	0.28	0.29	0.28
construction–flat–crosswalk-plain	0.43	0.33	0.35	0.40	0.26	0.31
construction–flat–sidewalk	0.51	0.49	0.48	0.32	0.29	0.30
marking–discrete–crosswalk-zebra	0.38	0.29	0.31	0.34	0.23	0.27
marking–discrete–stop-line	0.22	0.06	0.09	0.30	0.08	0.13
object–street-light	0.75	0.35	0.45	0.69	0.31	0.43
object–support–utility-pole	0.46	0.51	0.49	0.42	0.43	0.43
object–traffic-light–general-upright	0.66	0.64	0.67	0.63	0.61	0.62
object–traffic-light–pedestrians	0.49	0.48	0.47	0.47	0.43	0.45
object–traffic-sign–front	0.62	0.46	0.51	0.58	0.41	0.48
object–traffic-sign–information-parking	0.33	0.27	0.20	0.31	0.23	0.26
object–vehicle–bicycle	0.45	0.46	0.43	0.39	0.35	0.37
object–vehicle–bus	0.37	0.65	0.60	0.40	0.65	0.50
object–vehicle–car	0.62	0.74	0.74	0.56	0.64	0.60

As shown in Table 5.4, the highest per-class scores were achieved for *traffic lights*, *cars*, and *utility poles*. This shows that large, clear, and common objects were detected most reliably. In contrast, *bike lanes*, *stop lines*, and fine surface markings reached much lower scores, since they are smaller, harder to see, and often hidden in street-level images.

Precision–Recall Curves

To assess the overall detection quality of the Street-Level Model, precision–recall (PR) curves were generated for all object classes. Each curve visualises the relationship between *precision* (the proportion of correct detections among all predictions) and *recall* (the proportion of correctly detected instances among all ground-truth objects) across different detection thresholds. Curves located closer to the upper-right corner represent a higher balance between precision and recall, while lower or steeper curves indicate a faster decline in precision as recall increases (Ultralytics 2025).

Figure 5.2 displays the PR curves for all classes in grey, with the mean curve across all classes shown in blue. The blue line therefore summarises the average precision–recall relationship of the model, providing an aggregated view of its overall detection behaviour. The mean AP_{50}, corresponding to the average area under these curves, reaches 0.238. This value reflects the combined precision and recall performance of the model when evaluated over all classes.

As illustrated by the blue mean curve, precision decreases gradually with increasing recall. For instance, at a recall value of approximately 0.4, the corresponding precision

is about 0.25. This means that when the model successfully detects around 40% of all existing objects, roughly one quarter of its predictions are correct. Such relationships exemplify the general trade-off between completeness and accuracy that characterises object detection models.

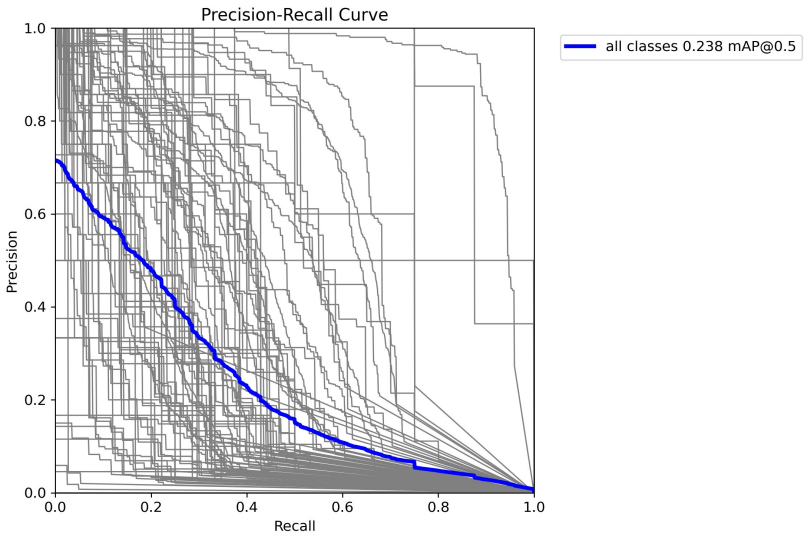


Figure 5.2: Precision–recall curves of the Street-Level Model. The blue line represents the mean performance across all object classes, while the grey lines correspond to the individual class curves.

Qualitative Examples

In addition to the quantitative evaluations, qualitative results were generated on street-level imagery from the Mapillary dataset. The Figures 5.3 - 5.5 illustrate three representative scenes from different urban locations in Zurich. The visualizations show detections of relevant traffic features such as vehicles, crosswalks, traffic lights, and markings directly on street-level photographs. These examples provide an impression of how the model performs when applied to crowd-sourced panoramic images with varying perspectives, lighting conditions, and levels of occlusion.

In Location 1 (Figure 5.3), a street segment with green surroundings and limited traffic is depicted. Location 2 (Figure 5.4) shows a clear and structured urban scene with visible markings and signage. Finally, Location 3 (Figure 5.5) presents a dense street-level view with multiple vehicles and objects in close proximity.

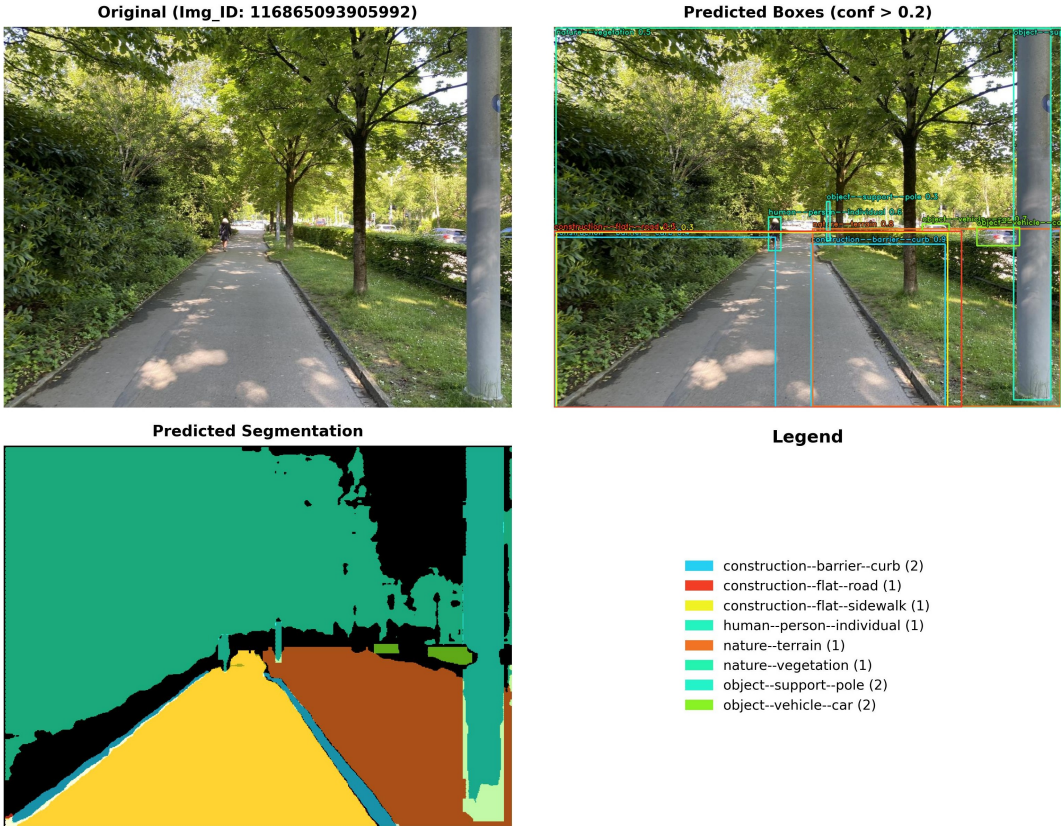


Figure 5.3: YOLO output example showing green surroundings (Location 1).

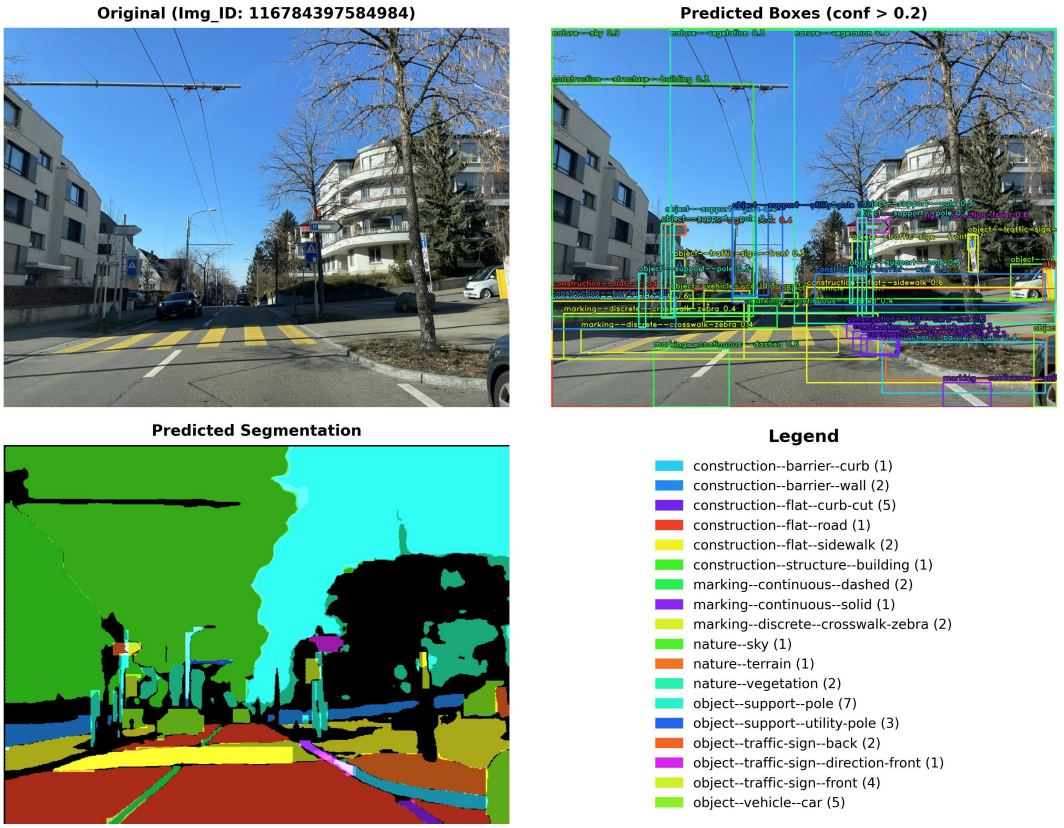


Figure 5.4: YOLO output example showing a clear and structured scene (Location 2).

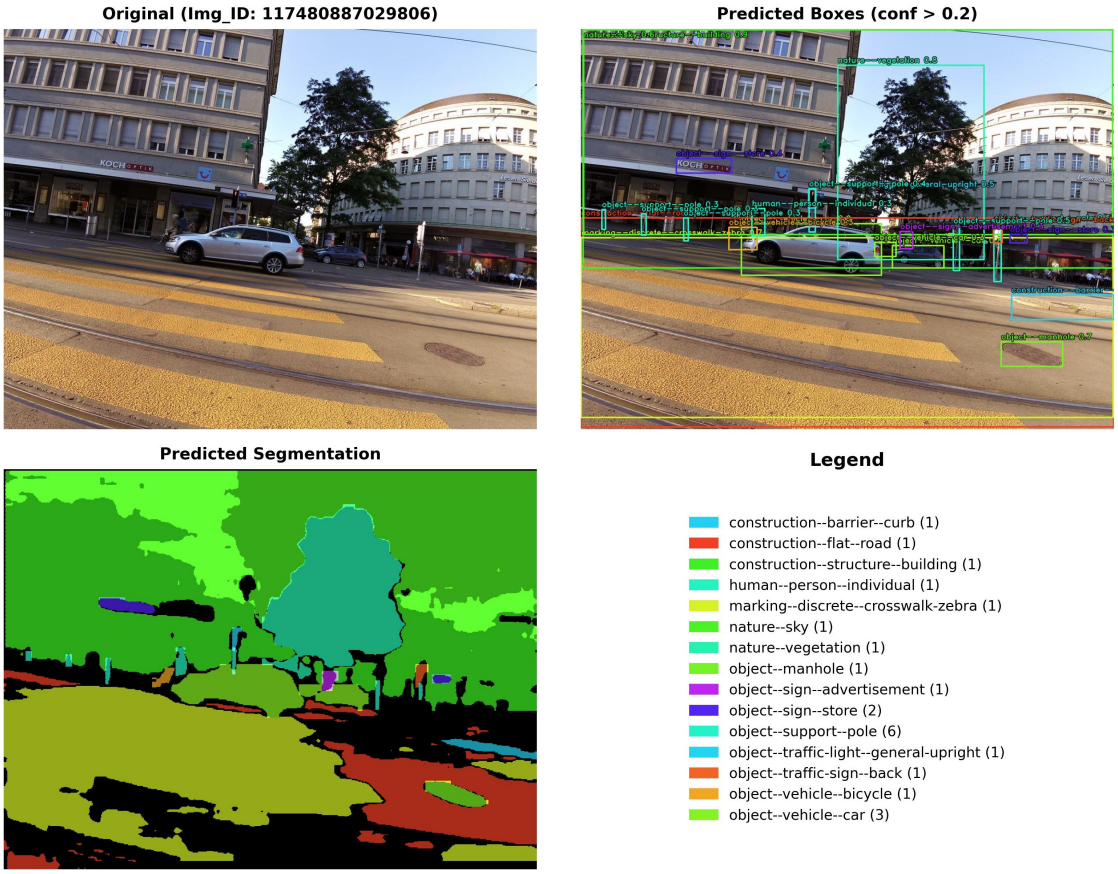


Figure 5.5: YOLO output example showing a dense street-level view (Location 3).

Runtime and Scalability

The four training runs correspond to the configurations described in Section 4.3.2. Table 5.5 summarises the training duration, architecture, and setup for each model. Model 1 (YOLO11m) was stopped after 36 epochs due to early stopping (patience = 5), requiring more than ten hours of training. Models 2 and 3 were subsequently fine-tuned based on the Street-level-model 1 for an additional 10 epochs each.

Table 5.5: Training and inference summary of Street-Level models.

Model	**Architecture**	**Epochs**	**Time [h]**	**Notes**
M 1	YOLO11m	36 (early stop)	∼10	Early stopping (patience = 5)
M 2	YOLO11m	10 (fine-tune)	∼11	Init. from Model 1
M 3	YOLO11m	10 (fine-tune)	∼12	Init. from Model 2
M 4	YOLO11s	100	∼9	Smaller tuned variant
Inference (Model 3)	—	—	∼30	∼1.2 Million images processed

During inference, all models were jointly applied to the complete Mapillary Vistas subset, processing approximately 1.2 million street-level images in just over 30 hours. This corresponds to an average throughput of roughly 40000 images per hour, or about

660 images per minute, using a single RTX 4080 GPU (16 GB VRAM) with Ultralytics YOLO11 and PyTorch. The resulting predictions comprised more than 17.9 million detected and georeferenced objects, which formed the spatial basis for the safety analysis.

5.2.2 Aerial Models

In addition to the street-level detections, aerial object detection was performed using SWISSIMAGE orthophotos (see Section 4.3.6).

Training and Overall Metrics

Four YOLO11 models were trained and evaluated on SWISSIMAGE orthophotos. These included one generalist model (Aerial Model 1) covering eight classes in Zurich and three specialised models (AM2–AM4) focusing on tram tracks or markings with training data from Zurich, Lucerne, and Bern.

Table 5.6 summarises the main performance metrics, model configurations, and computation parameters. Each model was trained and validated on SWISSIMAGE orthophotos using city-specific datasets, as described in Section 4.3.6. All models achieved high precision and recall values above 0.90. Model AM3 reached the highest mAP_{50} of 0.98 and $mAP_{50–95}$ of 0.93, while Models AM2 and AM4 both achieved mAP_{50} scores of 0.96. Model AM1 obtained slightly lower values of 0.93 (mAP_{50}) and 0.87 ($mAP_{50–95}$). Across all four models, a total of approximately 220,000 individual objects were detected, including about 14,700 road markings, 45,000 tram tracks, and further classes such as vehicles, crossings, and directional arrows.

Table 5.6: Performance and computation summary of YOLO11 models trained on SWISSIMAGE orthophotos (last epoch).

Property	**Aerial Model 1**	**Aerial Model 2**	**Aerial Model 3**	**Aerial Model 4**
Architecture	YOLO11X	YOLO11X	YOLO11X	YOLO11L
Epochs	100	70	100	100
Classes	8 (mixed)	1 (tram tracks)	9 (road markings)	1 (tram tracks)
Precision (B)	0.98	0.97	0.99	0.96
Recall (B)	0.90	0.93	0.95	0.91
mAP_{50} **(B)**	0.93	0.96	0.98	0.96
$mAP_{50–95}$ **(B)**	0.87	0.89	0.93	0.82
val/box_loss	0.31	0.42	0.32	0.61
Output	∼142,000 elements	∼45,000 tram tracks	∼14,700 markings	∼18,900 tram tracks

In addition to the quantitative metrics presented in Table 5.6, the prediction con-

fidences were analysed to evaluate the stability of class-wise detections. Figures 5.6 and 5.7 summarise the confidence distribution per class. Distinct variations are evident across the classes: *30 km/h zone marking* shows a median confidence of approximately 0.8, while *30 Zone* reaches around 0.9. High median confidences with comparatively narrow interquartile ranges are observed for *car* and *30 Zone*. In contrast, most other classes exhibit considerably larger interquartile ranges, indicating a broader variability in detection confidence. Classes such as *school zone marking*, *stop line marking*, *tram*, and *tram track* show intermediate median values, whereas the lowest medians (around 0.3) occur for *bicycle lane*, *bicycle marking*, *bus stop marking*, and *train*. The histograms in Figure 5.7 reveals a non-normal, U-shaped distribution across most classes, indicating that predictions tend to cluster either at low or high confidence levels, while intermediate confidence values are relatively infrequent. This pattern suggests a bimodal and non-normal distribution of prediction confidence, reflecting a clear separation between confident and uncertain detections.

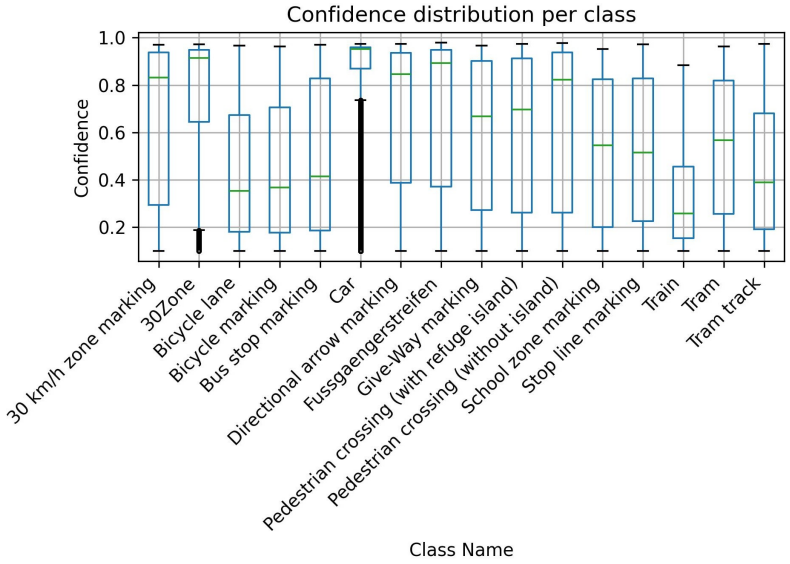


Figure 5.6: Boxplots showing the distribution of pred. confidence values for each class.

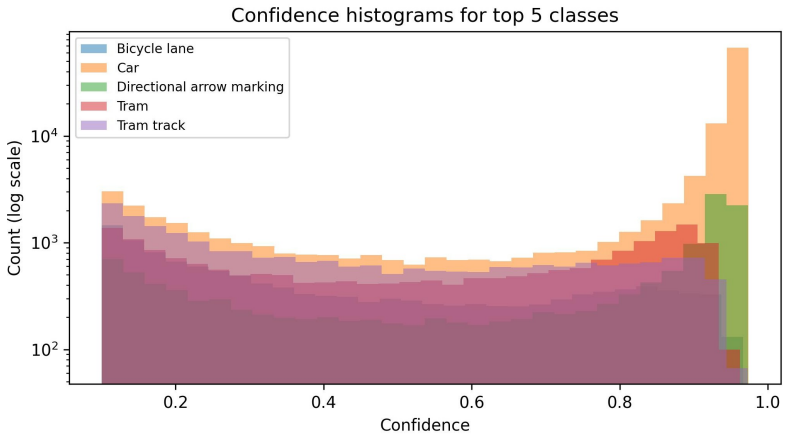
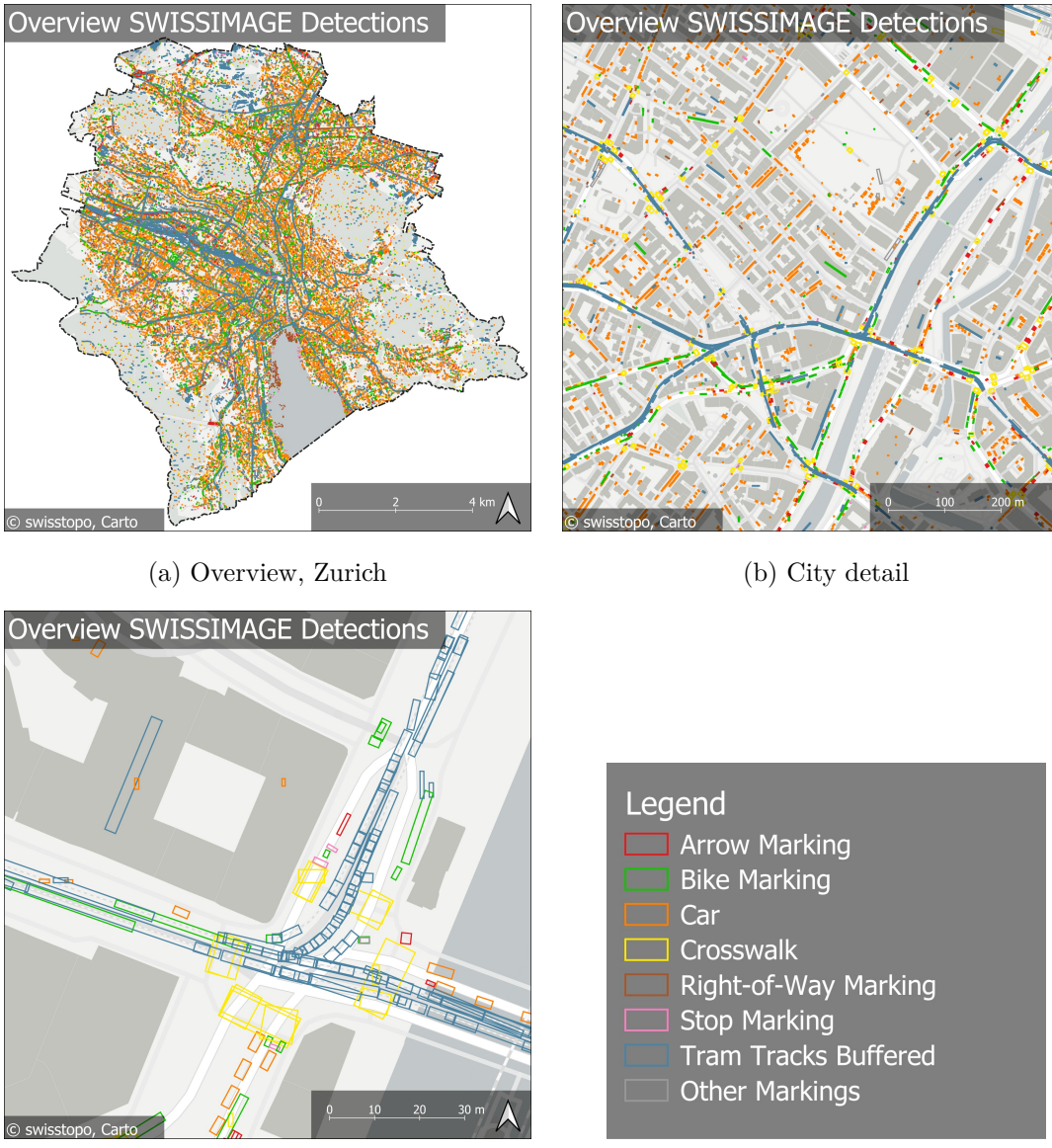


Figure 5.7: Confidence histograms for the five most frequent classes. The y-axis is displayed on a logarithmic scale.

Qualitative Examples

Figure 5.8 provides an overview of all aerial detections generated by models AM1–AM4. The visualisations show the combined GeoPackage containing all predictions at different spatial scales, from a citywide overview to detailed intersection views in Zurich. Subfigures (a)–(c) illustrate the spatial extent of the detections from city to intersection level, while subfigure (d) shows the legend used for visualising the combined outputs.



(a) Overview, Zurich

(b) City detail

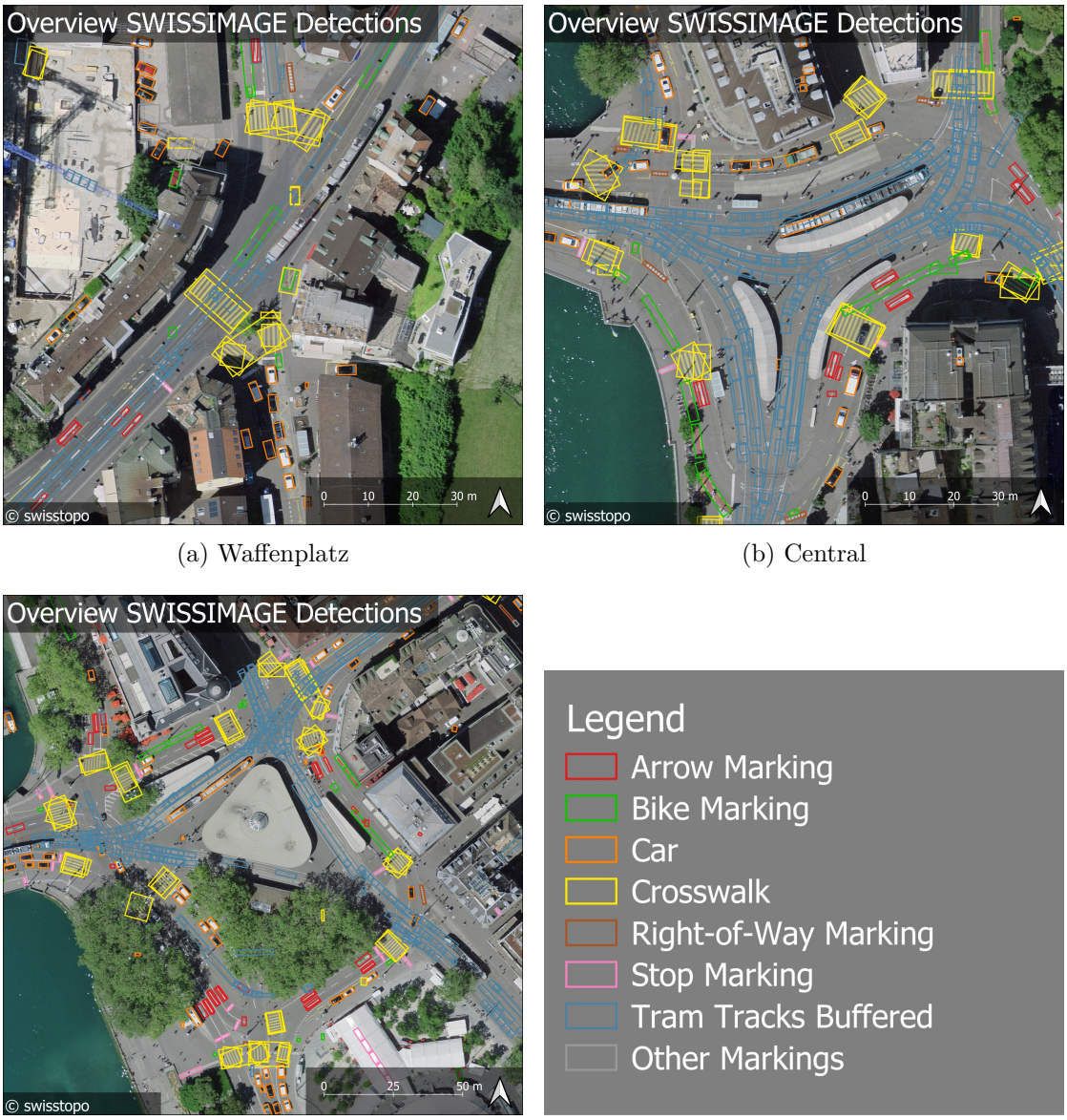
(c) Intersection detail (Stauffacherquai)

(d) Legend for combined predictions

Figure 5.8: Overview of the combined GeoPackage containing all predictions from AM1–AM4, illustrated at different spatial scales.

To illustrate the detection results in more detail, Figure 5.9 presents selected examples from the aerial inference results of models AM1–AM4. The subfigures show predicted objects such as tram tracks, road markings, crossings, and vehicles overlaid on SWISSIMAGE orthophotos. In Waffenplatz (Figure 5.9a), tram infrastructure and parallel

lane markings are clearly identified. At Central (Figure 5.9b), multiple vehicles and intersections are detected in a dense urban environment, while the Bellevue scene (Figure 5.9c) illustrates model performance under complex traffic conditions and variable surface contrasts. The legend for colour-coded classes is shown in subfigure (d).



(a) Waffenplatz (b) Central

(c) Bellevue (d) Legend

Figure 5.9: Qualitative examples of aerial YOLO11 predictions (AM1–AM4).

Runtime

Table 5.7 summarises the total training and inference runtimes of all aerial YOLO11 models. Training required between two and four hours per model, depending on architecture and dataset size, resulting in a combined training duration of approximately 11 hours. Inference across the full SWISSIMAGE coverage of Zurich (87.88 km^2) (Statistik Stadt Zürich 2025) took around 18 hours in total, corresponding to an average processing speed of roughly 3–4 minutes per km^2.

Table 5.7: Runtime summary of all aerial models on SWISSIMAGE orthophotos.

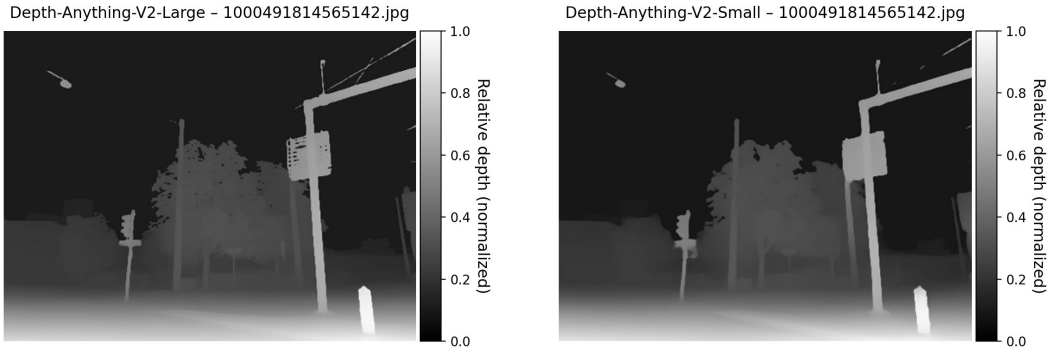
Property	**AM1**	**AM2**	**AM3**	**AM4**
Architecture	YOLO11X	YOLO11X	YOLO11X	YOLO11L
Epochs	100	70	100	100
Classes	8 (mixed)	1 (tram tracks)	9 (road markings)	1 (tram tracks)
Training [h]	3.0	2.0	4.3	2.0
Inference [h]	5.0	4.0	5.5	3.5
Time per km^2 [min]	3.4	2.7	3.8	2.4
Total Training	∼11.3 h total training time across all models			
Total Inference	∼18.0 h total inference time covering 87.88 km^2 (City of Zurich)			

5.3 Depth Estimation

5.3.1 Depth Estimation Results

Before the object detections from street-level and aerial imagery could be geographically interpreted, it was necessary to estimate the underlying scene geometry. Depth information allows detected objects to be placed not only in the image plane, but also in their approximate spatial context relative to the camera. This section therefore presents the results of the depth estimation pipeline introduced in Section 4.3.3.

The pipeline processed more than 1.2 million Mapillary street-level images using a self-supervised monocular depth approach. Two configurations of the *Depth-Anything V2* model were evaluated, differing in model size and inference speed. A comparison of the two different sizes of the model are visualized in Figure 5.10.



(a) Large model output (b) Small model output

Figure 5.10: Comparison of depth estimation outputs using the large and small model variants.

Figure 5.11 illustrates representative examples of the resulting depth maps across different urban settings in Zurich.

Figure 5.11: Examples of depth estimation outputs on different scenes: original images (left) and corresponding depth maps (right).

Runtime

The small model reached a throughput of approximately 47 images per second, allowing the full dataset of over 1.2 million images to be processed in about seven hours of inference time. Including validation, downscaling, and compression, the total runtime of the complete pipeline was approximately twelve hours.

Table 5.8: Benchmark results for depth estimation models on RTX 4080 GPU.

Model	**Throughput (images/s)**	**Total runtime for dataset**
Depth-Anything-V2-Large	17	∼19.6 h
Depth-Anything-V2-Small	47	∼7.1 h

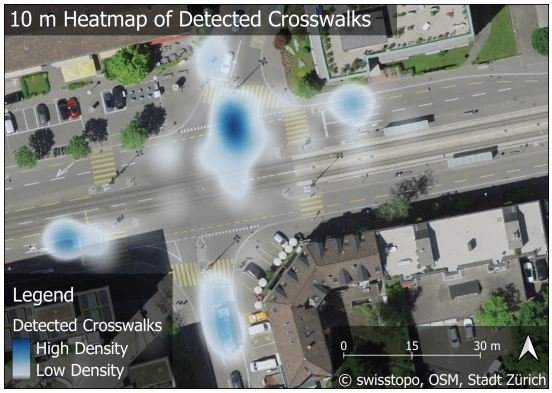
5.4 Object Geolocation Results

The depth maps produced in the previous step were integrated into the geolocation pipeline to spatially project all detected objects from image coordinates into geographic space. This section presents the results of the geolocation workflow (see Section 4.3.5).

Figures 5.12a–d illustrate this process for pedestrian crossings. In the left column, detections are still located at the camera position, while the right column shows the adjusted geolocations after applying the azimuth and distance shift.

(a) Original detection

(b) After geolocation shift



(c) Original detection

(d) After geolocation shift

Figure 5.12: Example of object georeferencing for pedestrian crossings.

Table 5.9 summarises the displacement values after the geolocation adjustment, reporting the ten classes with the highest and the five with the lowest mean distances. The largest mean displacements were observed for person groups, vehicle groups, and parking information signs, all exceeding 10 m on average. Vegetation, street lights, and buildings also showed relatively high values, corresponding to their frequent occurrence at greater distances from the camera. Smaller or near-camera objects such as traffic cones, temporary supports, and barriers typically showed mean displacements around 5 m. Across all categories, the median values were close to the predefined offset distances of 5 m and 7.5 m.

Table 5.9: Average displacement per object class (in meters). Top 10 classes with the highest mean displacement (above) and bottom 5 classes with the lowest mean displacement (below).

Class	**Count**	**Mean**	**Median**	**Std**	**Min**	**Max**
Pedestrians (groups)	1,828	12.10	7.50	6.44	5.00	20.00
Vehicles (groups)	25,000	11.79	7.50	6.40	0.00	20.00
Parking signs	22,552	10.74	7.50	6.33	0.00	20.00
Vegetation	1,770,751	10.74	7.50	6.17	0.00	20.00
Street lights	137,799	10.58	7.50	6.17	0.00	20.00
Buildings	867,221	10.46	7.50	6.04	0.00	20.00
Poles (groups)	312	10.20	7.50	5.74	5.00	20.00
Trams / rail vehicles	9,630	10.16	7.50	6.09	5.00	20.00
Pedestrians (individuals)	453,960	10.13	7.50	6.02	0.00	20.00
Cyclists	62,139	9.73	7.50	5.85	0.00	20.00
Bike lanes	13,305	5.57	5.00	2.12	5.00	20.00
Pedestrian areas	15,788	5.70	5.00	2.35	0.00	20.00
Crosswalks	2,327	5.72	5.00	1.77	5.00	20.00
Parking lots	45,574	5.93	5.00	2.69	0.00	20.00
Parking aisles	1	5.00	5.00	–	5.00	5.00

In total, more than 17.9 million detections were processed and exported in under two minutes.

5.5 Safety Classification

This section describes how a spatially explicit safety score was computed from the detected infrastructure features. The detected objects from both street-level and aerial models were georeferenced and mapped onto the pedestrian and school route network of Zurich. For each network segment, feature counts and weights were aggregated to calculate a standardized safety score. The results show how different areas of the city vary in terms of safety for schoolchildren.

5.5.1 Distribution of Safety Scores

This section presents the distribution of safety scores across all segments and districts. Tables provide summary statistics, while figures complement them with graphical and spatial representations. An additional analysis of SHAP-derived feature importance is included to identify the most influential variables in the ML model.

Overall distribution

Table 5.10 reports the summary statistics of safety scores for both methods across all segments. Both approaches cover the full range of possible scores from low to high values. The ML method reaches a mean of 86.3 with a median of 93.6, while the rule-based method achieves a very similar mean of 86.2 and a median of 92.4. Standard deviations of 16.9 and 17.7 indicate comparable overall variability.

Table 5.10: Summary statistics of safety scores across all segments, per method.

Method	Count	Min	Max	Mean	Median	Std.Dev
ML	102,590	1.1	100.0	86.3	93.6	16.9
Rule-based	102,590	0.0	92.4	86.2	92.4	17.7

Figure 5.13 displays the distribution of scores as density plots. The rule-based method produces a narrow peak around the mid-80s, whereas the ML method shows a broader distribution shifted towards higher values.

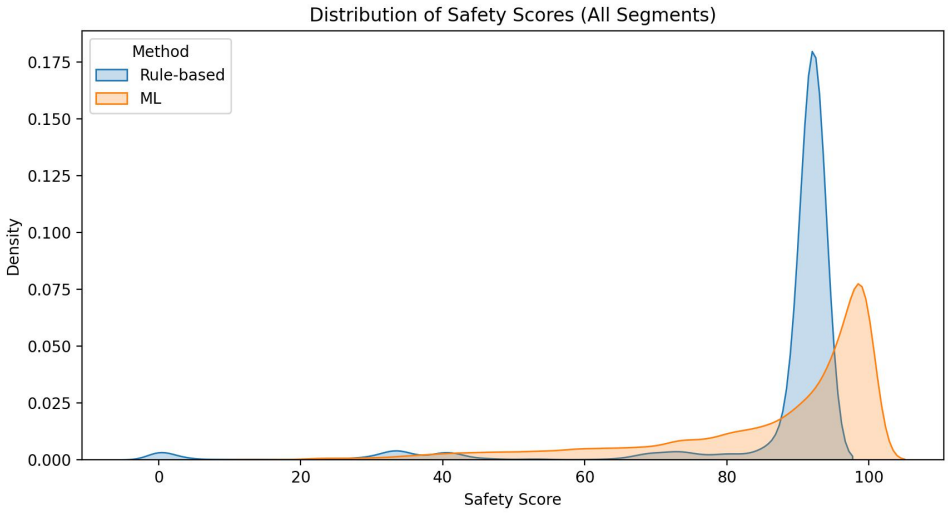


Figure 5.13: Distribution of safety scores across all segments, shown as density plots for both methods.

Spatial distribution

The spatial distribution of the resulting safety scores is illustrated on the following two A3 pages.

Predicted School Route Safety

Machine Learning Approach for the City of Zurich

Predicted safety values (0–100) for Zurich's school routes were estimated using a machine learning approach based on visual information from street-level and aerial imagery.

Objects such as crossings, sidewalks, and traffic lights were automatically identified through object detection and used to predict how safe each route segment is for schoolchildren.
Higher scores indicate safer walking routes.

1:50'000

District 11
District 10
District 12
District 6
District 5
District 9
District 4
District 1
District 7
District 3
District 8
District 2

Detailed View: District 1 (Altstadt)

1:15'000

Zurich HB
Central
Old Botanical Garden
Bellevue

Legend

Safety score (0–100)

- 0 - 20
- <20 - 50
- <50 - 60
- <60 - 70
- <70 - 80
- <80 - 90
- <90 - 100
- City of Zurich
- Districts of Zurich

Author: Claude Widmer
SWISSIMAGE (June 2025)
Mapillary imagery (June 2025)
Model: Random Forest with SMOTE
Sources: City of Zurich, Carto, Mapillary, swisstopo

Predicted School Route Safety

Rule-based Approach for the City of Zurich

Predicted safety values (0–100) for Zurich's school routes were derived from a rule-based scoring system using detected features from street-level and aerial imagery.

Each feature type, such as crossings, sidewalks, and traffic lights, was assigned predefined weights and thresholds to estimate how safe each route segment is for schoolchildren.
Higher scores indicate safer walking routes.

1:50'000

District 11
District 10
District 12
District 6
District 5
District 9
District 4
District 1
District 7
District 3
District 8
District 2

Detailed View: District 1 (Altstadt)

1:15'000

Zurich HB
Central
Old Botanical Garden
Bellevue

Legend

Safety score (0–100)

- 0 - 20
- <20 - 50
- <50 - 60
- <60 - 70
- <70 - 80
- <80 - 90
- <90 - 100
- City of Zurich
- Districts of Zurich

Author: Claude Widmer
SWISSIMAGE (June 2025)
Mapillary imagery (June 2025)
Model: Rule-Based model with weights
Sources: City of Zurich, Carto, Mapillary, swisstopo

To highlight differences between the two approaches, Figure 5.14 shows the segment-wise subtraction of rule-based scores from ML-based scores. Positive values (blue tones) indicate higher scores under the ML method, while negative values (red tones) represent higher scores under the rule-based method.

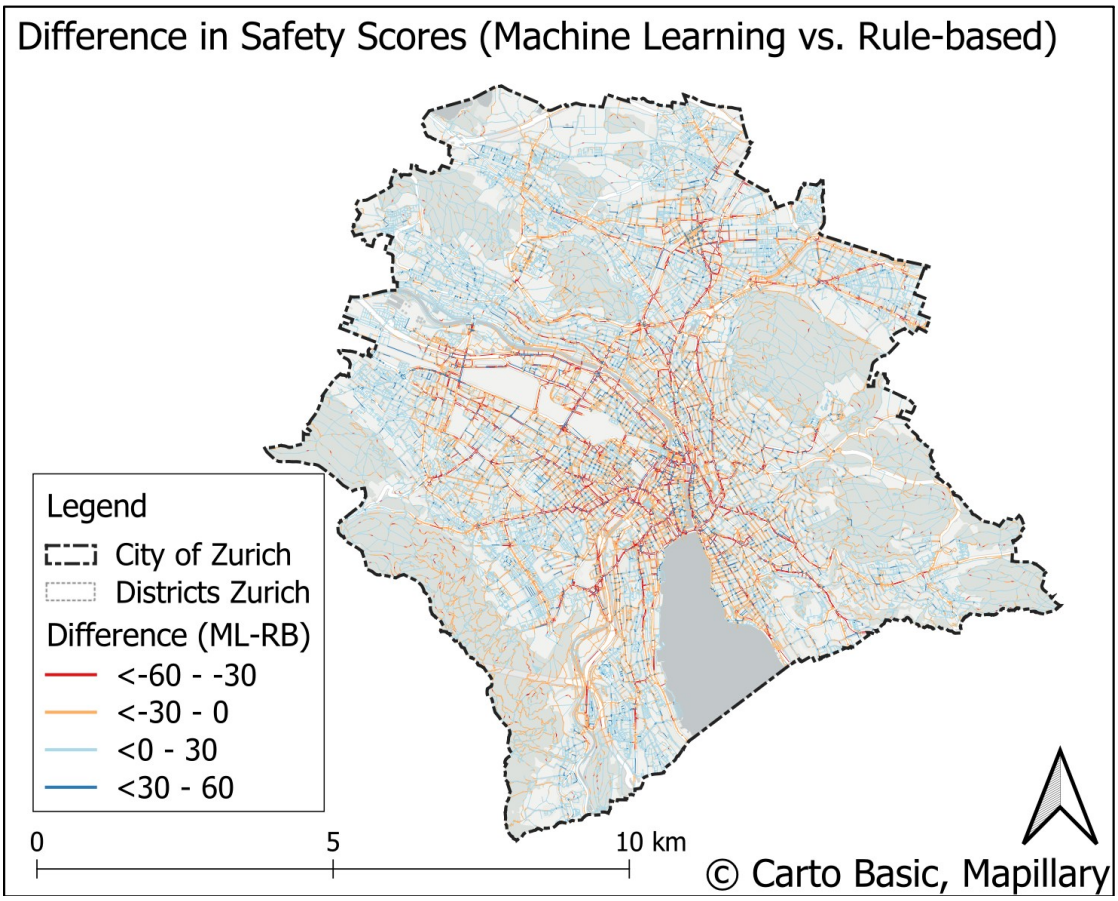


Figure 5.14: Difference in safety scores (ML minus rule-based) across all network segments. Positive values (blue) indicate segments with higher ML scores, negative values (red) indicate higher rule-based scores.

Feature Importance of the ML Classifier

To interpret which input variables contributed most to the predicted safety scores, global SHAP values were computed across all classes (see Section 4.4.3). Figure 5.15 shows the mean absolute SHAP values of the 30 most important features. Dashed road markings and traffic lights exhibit the strongest positive influence, followed by tram-related and pedestrian-crossing features. The complete list of features and their corresponding SHAP values is provided in the Appendix.

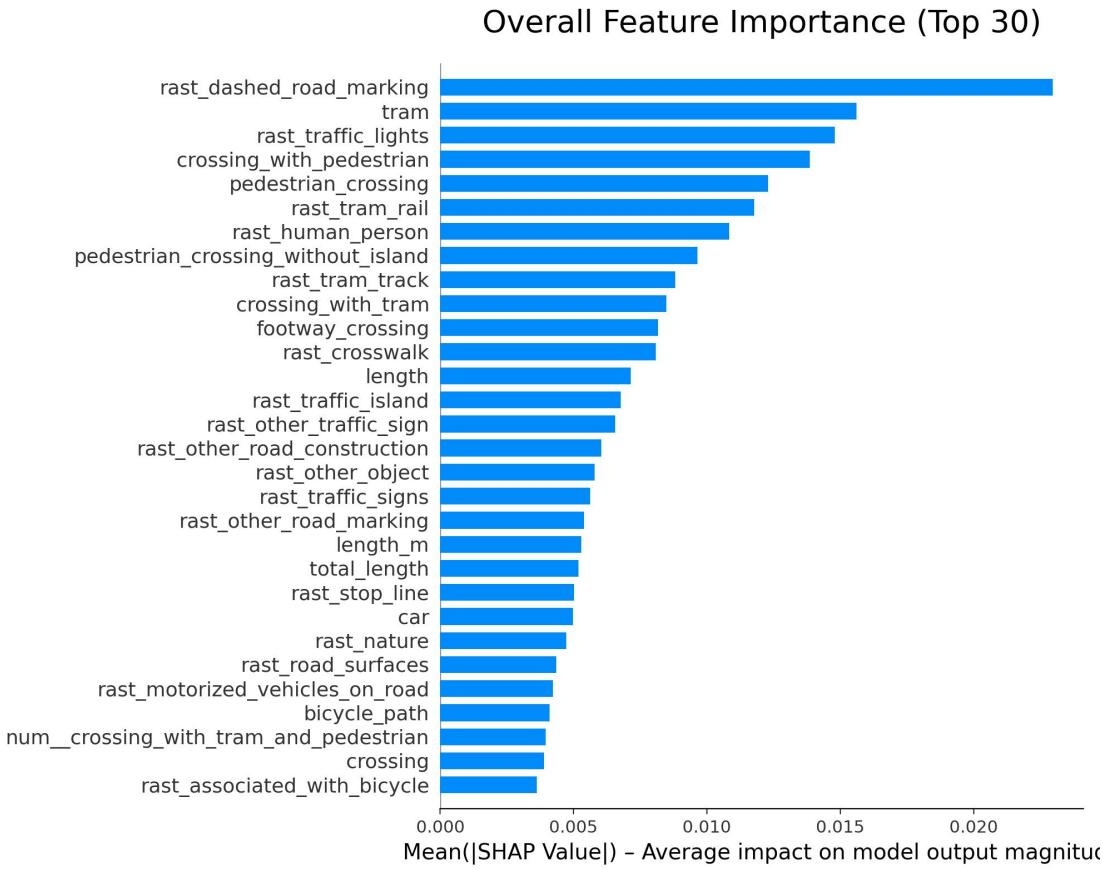


Figure 5.15: Overall feature importance of the ML classifier based on mean absolute SHAP values.

District-level distribution

Table 5.11 summarises the safety scores per district (Stadtkreis) for both classification methods. The ML approach shows mean values ranging from about 70.7 in central districts (Kreis 1) up to over 91.2 in outer districts such as Kreis 10. The rule-based results follow a similar pattern, with slightly lower mean values across all districts.

Table 5.11: Summary statistics of safety scores per district (Stadtkreis), for ML and rule-based methods.

District	Min	Max	Mean	Median	Std.Dev
1	1.8	100.0	70.7	75.5	22.0
2	2.8	100.0	88.4	94.9	15.3
3	2.6	100.0	87.4	94.2	15.5
4	1.7	100.0	82.5	89.6	17.2
5	1.4	100.0	83.8	91.3	16.7
6	3.7	100.0	87.7	94.1	15.0
7	3.2	100.0	89.1	94.9	14.3
8	1.2	100.0	83.8	90.9	16.9
9	2.2	100.0	88.1	94.4	14.8
10	9.3	100.0	91.2	96.2	12.1
11	1.5	100.0	89.6	96.5	15.2
12	4.5	100.0	88.8	95.9	15.1

(a) ML method

District	Min	Max	Mean	Median	Std.Dev
1	0.0	88.2	69.9	76.8	21.9
2	0.0	92.4	86.3	92.4	17.7
3	0.0	92.1	85.2	91.4	18.0
4	0.0	90.0	81.1	87.8	19.0
5	0.0	89.3	82.3	89.2	18.4
6	0.0	91.4	85.1	91.0	17.3
7	0.0	90.0	86.0	91.8	16.6
8	0.0	89.1	83.7	89.3	18.2
9	0.0	88.1	84.6	90.5	17.6
10	0.0	86.5	88.3	92.1	16.4
11	0.0	89.7	87.0	92.0	17.0
12	0.0	88.3	86.5	91.7	17.2

(b) Rule-based method

Figure 5.16 shows the distribution of safety scores per district for both methods. In the ML-based results, central districts such as Kreis 1 exhibit a wider spread and lower median values compared to the outer districts, where scores are generally higher and more consistent.

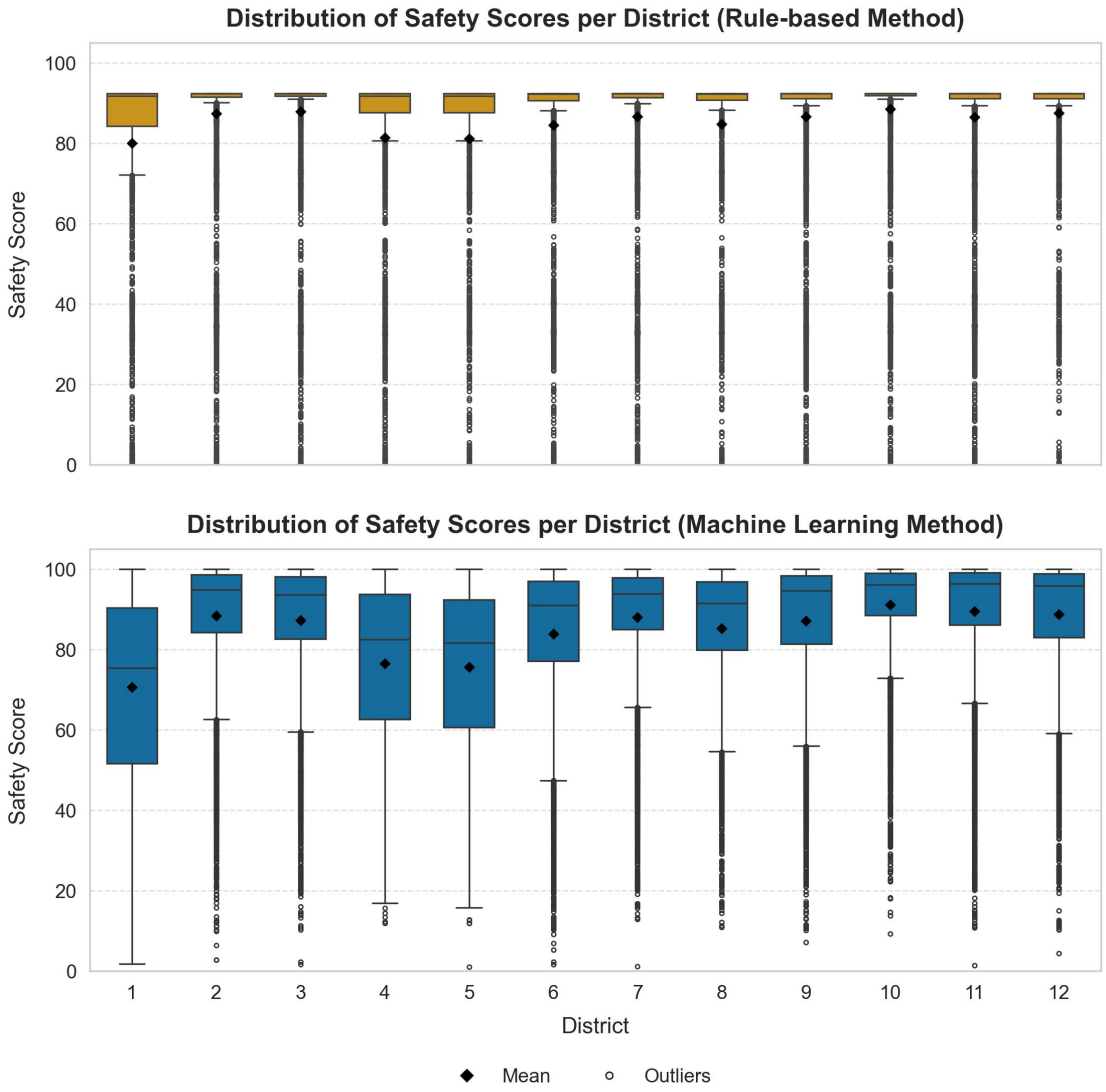


Figure 5.16: Boxplots of safety scores per district for both methods.

The spatial patterns of mean district-level safety scores are shown in Figures 5.17 and 5.18. Figure 5.17 displays the results from the ML-based method, while Figure 5.18 shows the corresponding rule-based results.

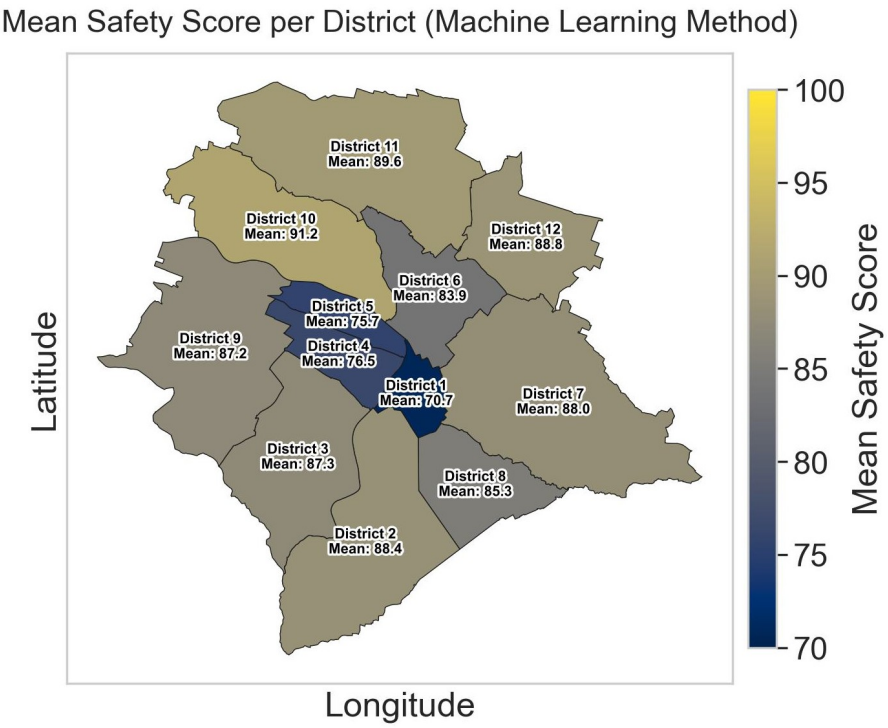


Figure 5.17: Mean safety score per district (ML method).

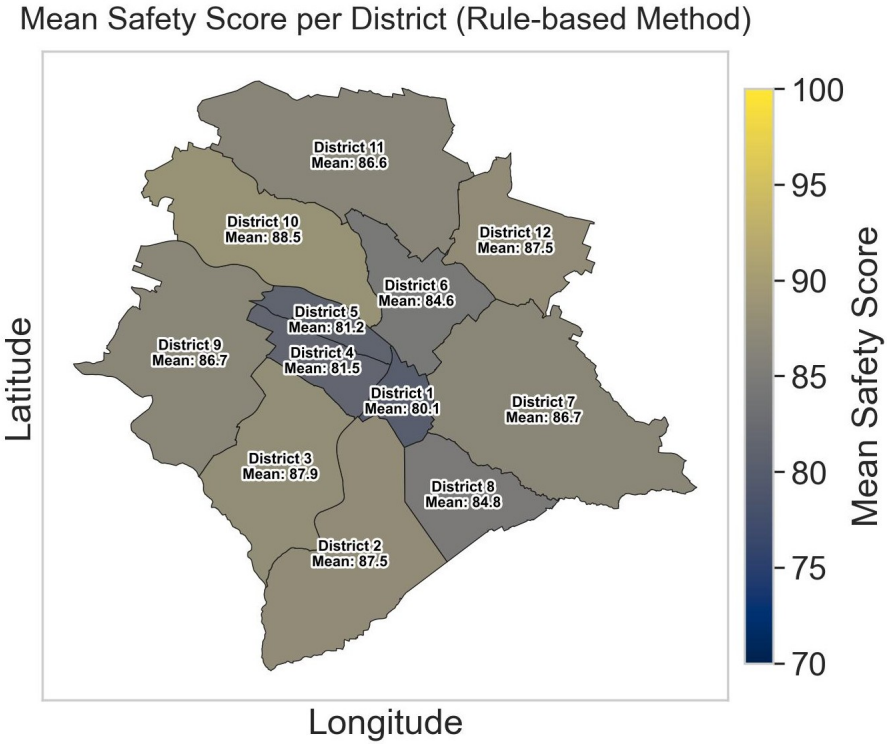


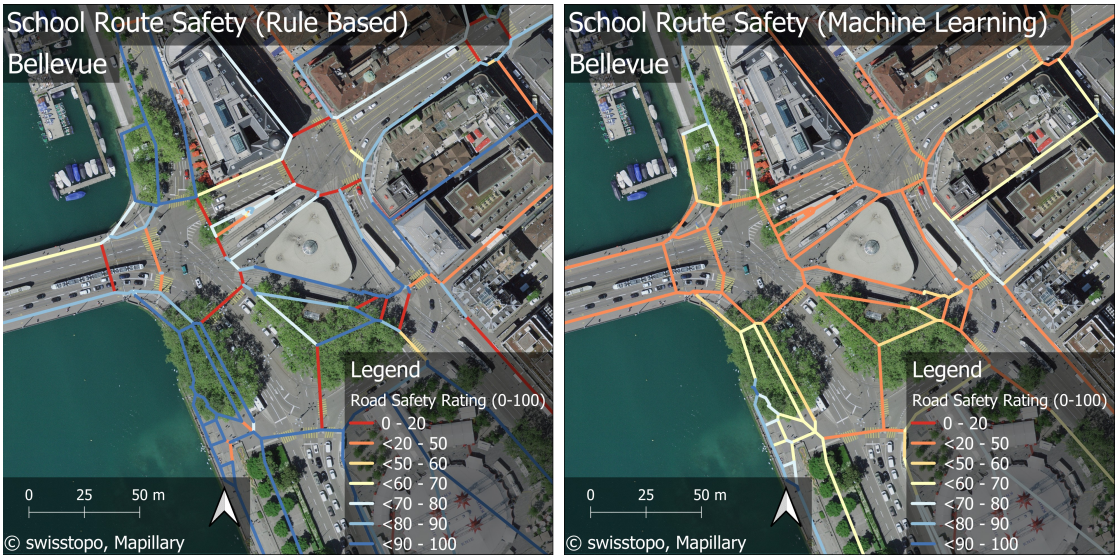
Figure 5.18: Mean safety score per district (rule-based method).

5.5.2 Case Examples

To provide a better understanding of the safety scores, four case examples are presented. Each example shows the same location with the rule-based method on the left and the

ML method on the right. This allows for a direct comparison of both approaches in different urban contexts.

Figure 5.19 shows the Bellevue area, one of Zurich's most prominent traffic hubs with multiple tram lines and high pedestrian activity. Both methods identify several low-scoring segments near the main crossing zones. The Machine-Learning (ML) method assigns generally lower safety scores across the entire intersection, particularly in areas close to tram tracks and complex crossing layouts. In contrast, the rule-based method produces distinctly lower ratings at tram-related crossings and intersections with pedestrian stripes, while assigning higher values along adjacent sidewalks and pedestrian corridors.



(a) Rule-based (b) ML

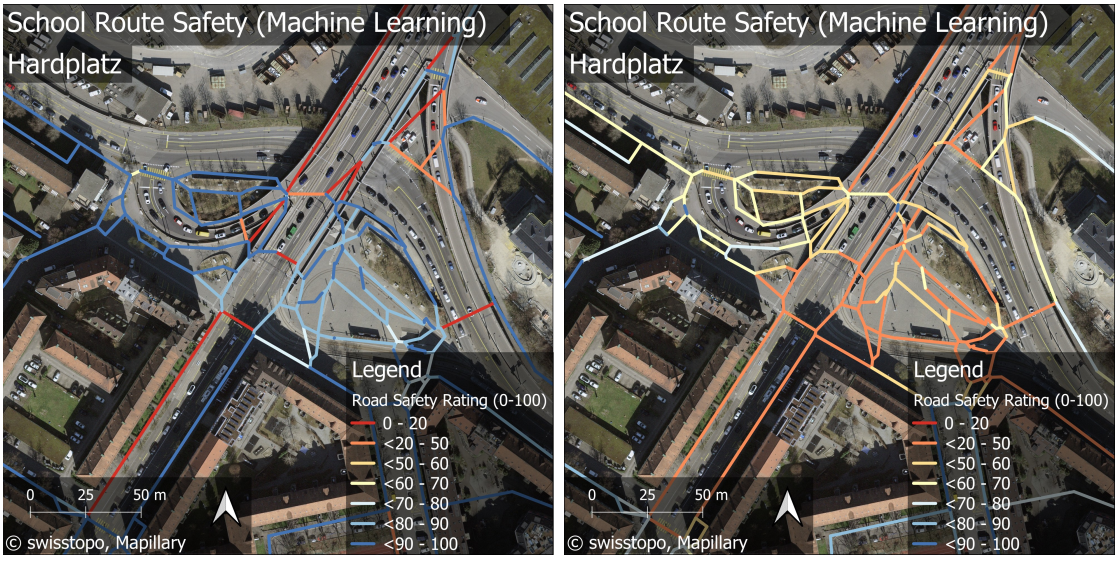
Figure 5.19: Case example Bellevue

Figure 5.20 shows the Central intersection. Both methods capture lower safety scores in the main crossing areas near the tram tracks and vehicle junctions.

(a) Rule-based (b) ML

Figure 5.20: Case example Central

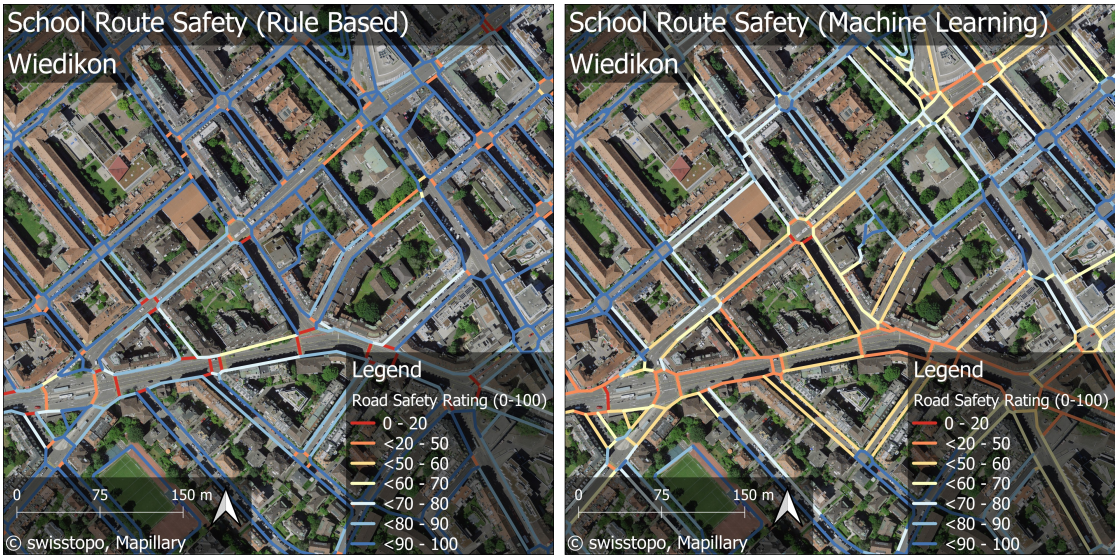
Figure 5.21 shows the Hardplatz area, a large transport interchange characterised by road and tram infrastructure. Both methods depict a mix of high and low safety scores. The ML method highlights corridor-like high-safety areas along main pedestrian routes, while the rule-based results are more fragmented around crossing zones.



(a) Rule-based (b) ML

Figure 5.21: Case example Hardplatz

Figure 5.22 shows an example of a residential area in Wiedikon with smaller streets and a regular urban grid. In this calmer neighbourhood, the machine-learning method generally assigns lower safety scores to the sidewalks. The rule-based method, however, assigns higher safety scores to the neighbourhood, but lower scores to the crossings.



(a) Rule-based (b) ML

Figure 5.22: Case example Wiedikon

5.6 Routing Outcomes: Safe vs. Shortest Paths

This section investigates how safety-aware routing changes the selected paths compared to purely distance-based routing. The analysis is conducted for several predefined *origin–destination (OD) pairs*, i.e. specific start–end locations within the study area. For each OD pair and for both networks (rule-based and ML), four route variants are computed based on the cost formulation from Section 4.4.5: (i) Fastest ($\alpha = 1$, $\beta = 0$), (ii) Balanced ($\alpha = 1$, $\beta = 1$), (iii) Safest ($\alpha = 1$, $\beta = 3$), and (iv) Absolute Safest ($\alpha = 0$, $\beta = 1$). All routes were generated through the internal API described in Section "3.5: Routing with Cost Function", ensuring full reproducibility and consistency across all routing scenarios.

5.6.1 Case Examples

Overview

Tables 5.12 and 5.13 present the four route variants for selected OD pairs, separately for the ML-based and rule-based networks. For each case, the tables report the total distance (m), the length-weighted mean safety score, and the relative changes with respect to the *Fastest* route.

For the ML-based network (Table 5.12), the *Fastest* routes serve as the baseline, with distances ranging from 1486 m (R3) to 7797 m (R2) and mean safety values between 65.9 (R1) and 78.1 (R2). The *Balanced* routes are slightly longer, for example increasing R1 from 2135 m to 2252 m (+5.5%) while the safety score rises from 65.9 to 79.4 (+20.5%). The *Absolute Safest* routes are substantially longer but yield the highest safety improvements. For instance, R1 increases from 2135 m to 2645 m (+23.9%), reaching a safety

score of 85.6 (+29.9%).

Table 5.12: Case examples per OD and method (ML network): Distances in meters; safety is length-weighted mean. Relative changes are measured vs. *Fastest*.

Route	Case	Dist. (m)	Safety	ΔDist (%)	ΔSafety (%)
R1	Fastest	2135	65.9	0.0	0.0
R1	Balanced	2252	79.4	5.5	20.5
R1	Safest	2347	82.7	9.9	25.6
R1	Absolute Safest	2645	85.6	23.9	29.9
R2	Fastest	7797	78.1	0.0	0.0
R2	Balanced	8008	86.1	2.7	10.3
R2	Safest	8141	87.2	4.4	11.6
R2	Absolute Safest	11814	93.2	51.5	19.3
R3	Fastest	1486	74.2	0.0	0.0
R3	Balanced	1534	83.3	3.2	12.2
R3	Safest	1541	84.2	3.7	13.5
R3	Absolute Safest	1805	88.9	21.5	19.8

For the rule-based network (Table 5.13), the *Fastest* routes form the baseline, with distances between 1486 m (R3) and 7797 m (R2) and mean safety scores between 82.7 and 83.4 (R3). The *Balanced* routes are only marginally longer, for example increasing R2 from 7797 m to 7834 m (+0.5%), accompanied by a moderate safety gain from 83.1 to 85.4 (+2.8%). The *Safest* routes follow a similar pattern, such as R1 increasing from 2135 m to 2299 m (+7.7%) while its safety score improves from 82.7 to 88.6 (+7.1%). The *Absolute Safest* routes remain close to the baseline in terms of distance, e.g. R3 extends from 1486 m to 1520 m (+2.3%) while safety improves slightly from 83.4 to 85.0 (+1.9%).

Table 5.13: Case examples per OD and method (Rule-based network): Distances in meters; safety is length-weighted mean. Relative changes are measured vs. *Fastest*.

Route	Case	Dist. (m)	Safety	ΔDist (%)	ΔSafety (%)
R1	Fastest	2135	82.7	0.0	0.0
R1	Balanced	2167	85.3	1.5	3.2
R1	Safest	2299	88.6	7.7	7.1
R1	Absolute Safest	2300	88.6	7.7	7.1
R2	Fastest	7797	83.1	0.0	0.0
R2	Balanced	7834	85.4	0.5	2.8
R2	Safest	7991	86.9	2.5	4.6
R2	Absolute Safest	9045	89.9	16.0	8.2
R3	Fastest	1486	83.4	0.0	0.0
R3	Balanced	1496	84.2	0.7	0.9
R3	Safest	1499	84.3	0.9	1.0
R3	Absolute Safest	1520	85.0	2.3	1.9

The following figures illustrate the spatial realisations of these routing outcomes for selected origin–destination pairs.

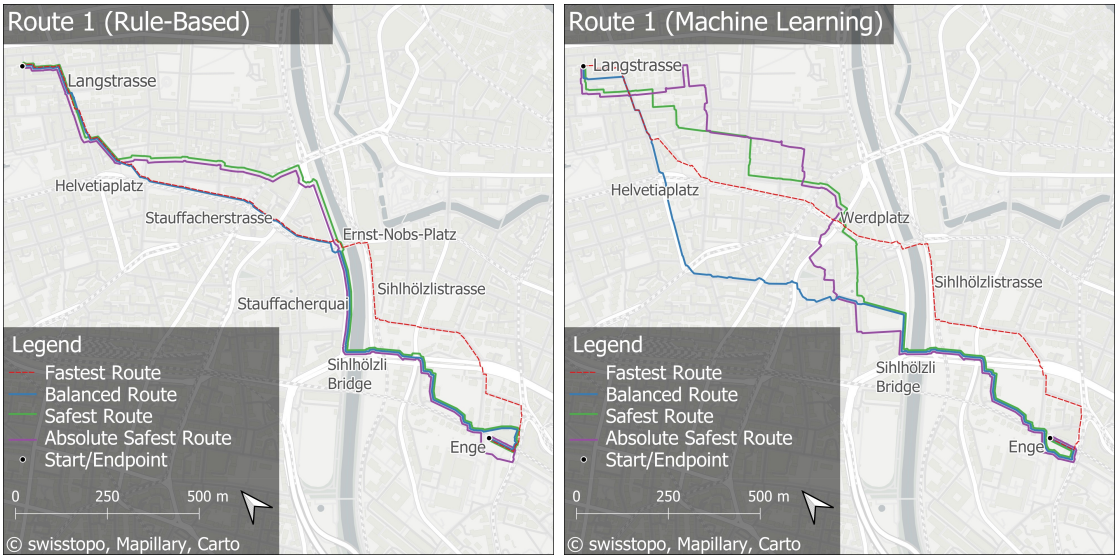
Map Examples

Figures 5.23–5.25 illustrate the four routing variants for each OD pair. Left panels show the rule-based network; right panels show the ML network. Since $\beta = 0$, the *Fastest* route is identical across both methods.

Figure 5.23 illustrates the four routing variants for **Route 1**, linking *Bahnhof Enge* with *Langstrasse* via the inner-western part of the city.

In the rule-based network (panel (a)), all routes begin at *Bahnhof Enge* and cross the Sihl via the *Sihlhölzli Bridge*. The *Fastest* route follows the main corridor along the *Stauffacherquai* and continues directly towards *Helvetiaplatz* and *Langstrasse*. Both the *Balanced* and *Safest* routes also use the *Sihlhölzli Bridge*, but leave the main corridor shortly after crossing the river, following a parallel side street with higher safety scores. The *Balanced* route rejoins the main corridor near *Helvetiaplatz* before reaching the destination, whereas the *Absolute Safest* route bypasses *Helvetiaplatz* entirely, passing through the *Werdplatz* area and continuing along smaller side streets to the endpoint at *Langstrasse*.

In the ML-based network (panel (b)), the general structure of the routes remains similar, but the model assigns lower safety scores along tram corridors and busy intersections. As a result, all variants avoid the *Sihlhölzistrasse* and prefer the crossing over the *Sihlhölzli Bridge*, with the *Safest* and *Absolute Safest* routes again diverting into quieter residential streets west of *Helvetiaplatz*.



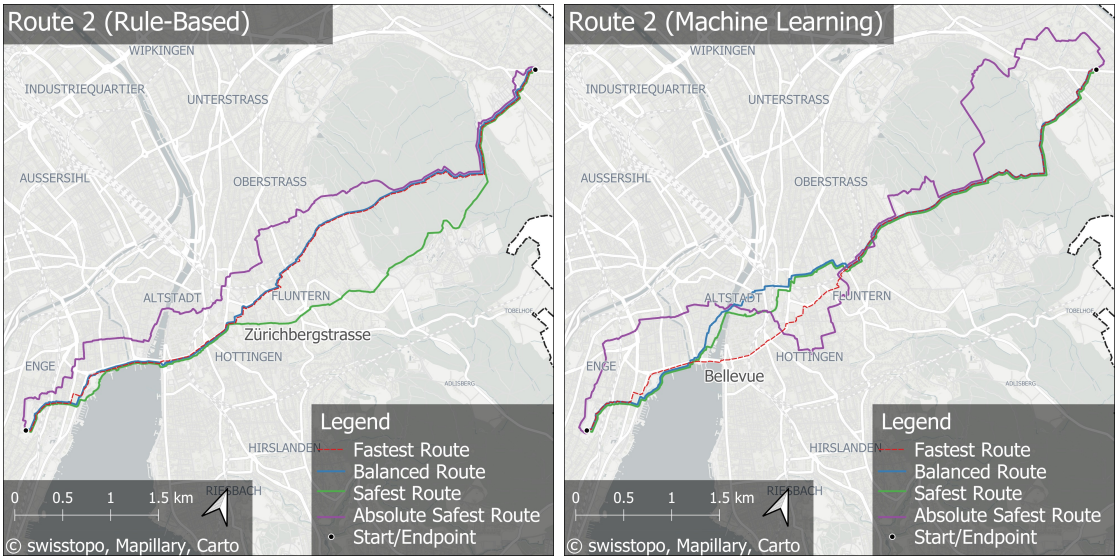
(a) Route 1 – Rule-based (b) Route 1 – ML

Figure 5.23: Routing example for Route 1: rule-based (left) vs. ML (right).

Figure 5.24 presents the routing variants for **Route 2**, representing a longer connection across the city from *Enge* on the lakeshore to *Schwamendingen* in the north. This case does not represent a typical school route but serves to illustrate the general routing behaviour of both models. It covers a larger spatial extent and includes several major intersections and elevation changes along the route.

In the rule-based network (panel (a)), both the *Fastest* and *Balanced* routes follow an almost identical alignment. They cross the city centre via *Bellevue*, an area characterised by heavy traffic and complex tram intersections, before continuing uphill through *Fluntern* towards the northern districts. The *Safest* variant follows a similar general direction but diverts after *Hottingen*, taking the *Zürichbergstrasse* and continuing along the forested ridge of the *Hardwald* towards *Schwamendingen*.

In the ML-based network (panel (b)), the routing behaviour differs more distinctly. Here, the routes avoid the *Bellevue* area entirely and instead traverse the *Altstadt* corridor. After passing *Fluntern*, all variants converge again and continue along a similar alignment to the *Fastest* route towards the destination in *Schwamendingen*.

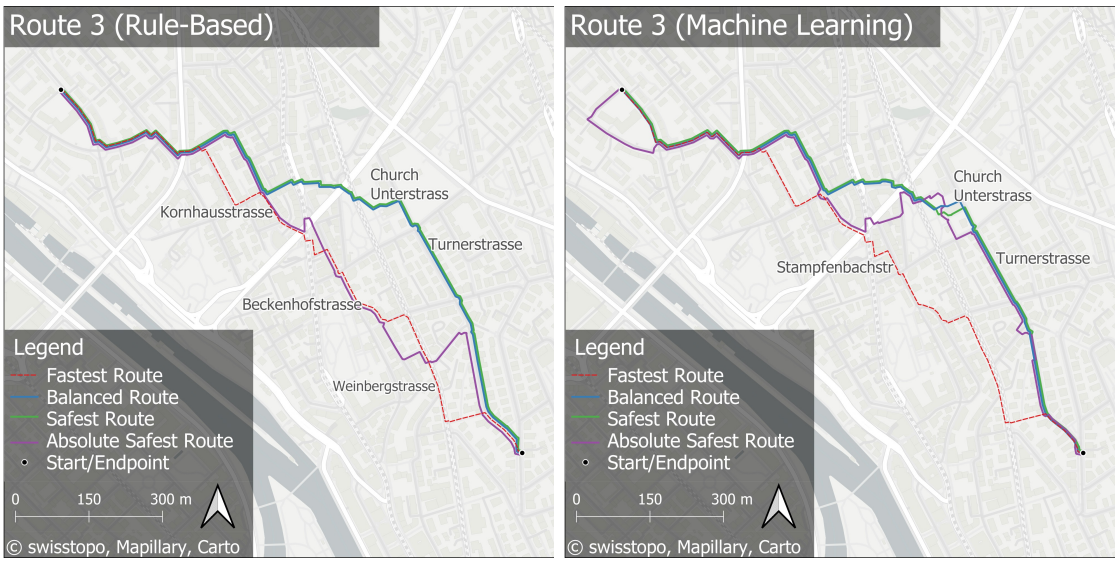


(a) Route 2 – Rule-based (b) Route 2 – ML

Figure 5.24: Routing example for Route 2: rule-based (left) vs. ML (right).

For **Route 3** (Figure 5.25), the rule-based network shows distinct but modest variations between the routing variants. The *Fastest* route follows the *Weinbergstrasse*, crossing one tram line, and continues via the *Beckenhofstrasse* towards the *Kornhausstrasse*, from where all routes share a similar alignment to the destination. Both the *Balanced* and *Safest* routes instead use the *Turnerstrasse*, which avoids tram tracks, then pass by the *Kirche Unterstrass* and cross the *Weinbergstrasse* once before rejoining a nearly identical alignment to the destination. The *Absolute Safest* route takes a more complex and longer detour via the *Beckenhofstrasse*, prioritising sections with the highest local safety scores.

In the ML-based network, the *Balanced* and *Safest* routes are nearly identical to their counterparts in the rule-based network, following a similar course via the *Turnerstrasse* and *Weinbergstrasse*. However, the *Absolute Safest* route deviates more noticeably in the area around the *Stampfenbachstrasse*, where it follows an alternative alignment to avoid lower-rated segments identified by the ML model.



(a) Route 3 – Rule-based (b) Route 3 – ML

Figure 5.25: Routing example for Route 3: rule-based (left) vs. ML (right).

6 Discussion

This chapter reflects on the findings of the thesis by examining the methods, results, and their implications. The aim is to interpret the outcomes in relation to existing literature, planning practices, and methodological considerations.The discussion is structured into five sections: detection and data basis (Section 6.1), safety scoring (Section 6.2), routing and practical relevance (Section 6.3), comparison with existing plans and stakeholder feedback (Section 6.4), and limitations (Section 6.6). Each section concludes with a box of key takeaways that summarises the main points.

6.1 Detection and Data Basis

Focus RQ1

Which safety-relevant infrastructure elements can be reliably detected, and how does performance differ between aerial imagery and systematic street-level data?

Research Question 1 addresses the foundation of the entire framework: the ability to automatically identify safety-relevant features in urban imagery. Reliable detection of crossings, sidewalks, traffic lights, and other key elements is a prerequisite for all subsequent stages of scoring and routing. This section therefore evaluates how effectively the trained YOLO models captured such infrastructure from both street-level and aerial perspectives, and how the two data sources complement each other in terms of coverage, precision, and detail.

The empirical results reveal substantial variation in detection performance depending on the type of element and the imagery source. Among the four trained YOLO11 variants for street-level data, Street-Level-Model 3 and 4 performed best, reaching overall mean average precision scores of 0.28 for bounding boxes (mAP50) and 0.23 for segmentation (mAP50) (see Table 5.2 and 5.3). Performance differs strongly across classes (Table 5.4). Traffic lights (upright) and cars are among the most reliably detected elements, with AP50 values of 0.67 and 0.74 and segmentation F1-scores of 0.62 and 0.60. Sidewalks and zebra crossings achieved medium results in the range of 0.30 to 0.48, while stop lines and bicycle lanes remained very weak, often below 0.15 AP50. The overall mAP50 of Street-Level-Model 3 across all classes was only 0.238, underscoring the gap between results on complex, real-world imagery and those obtained on benchmark datasets.

Looking at related work, previous studies have reported much higher accuracies when detection tasks are constrained to specific, visually distinctive features under controlled conditions. Kaya et al. (2023) trained a YOLOv7 model for pedestrian crosswalk detection and reported accuracies of up to 98%, demonstrating that such large and visually dominant features can be recognised with almost perfect reliability under curated con-

ditions. At the same time, their study also noted that smaller variations in crosswalk design and environmental conditions reduce robustness, which is consistent with the difficulties observed in our results for more subtle or small-scale elements.

The second focus was on the aerial models (see Section 5.2.2), which were trained on orthophotos. The aerial-trained models achieved substantially higher values (Table 5.6). Areal Model 3 (AM3), trained specifically on markings in Lucerne, reached 0.98 mAP50 and 0.93 mAP50–95, with precision and recall close to unity (0.99 and 0.95 respectively). Even the more generalist Aerial Model 1 (AM1) achieved 0.93 mAP50, while tram-focused models AM2 and AM4 scored 0.96 mAP50. These results demonstrate that aerial imagery is highly suited for detecting large-scale, structural features such as road markings, tram tracks, and intersections, with very low misclassification rates (Figure 5.9). Similar observations have been made by Ding et al. (2021), who introduced the large-scale DOTA benchmark for object detection in aerial images. In their study, models trained on aerial views also achieved strong accuracy despite challenges such as object orientation, varying densities, and scale variation. This supports the finding that aerial imagery is particularly well suited for the reliable detection of structural features, even under complex visual conditions. Thus, aerial imagery excels at structural features at the network level, while street-level views remain indispensable for pedestrian-scale details.

While per-image performance of crowdsourced street-level data is weaker than that of controlled datasets, the thesis processed more than 1.2 million Mapillary images. This large sample size helped stabilise the aggregated results: individual errors and noise were effectively averaged out across the dataset. The effect can be understood in analogy to the Central Limit Theorem, which implies that the variance of aggregated estimates tends to decrease as the number of observations increases (Islam 2018). In practice, this means that although single Mapillary images often contain artefacts such as blur, occlusion, or poor lighting, the overall city-wide detection outcomes become more robust when aggregated at scale. Similar observations have been made in comparative studies on Mapillary Vistas, where YOLOv7 outperformed other detectors but benefitted most from the dataset's size and heterogeneity (Z. Yang et al. 2022). This illustrates that quantity and coverage can partly compensate for the lower quality of crowdsourced imagery, producing meaningful spatial patterns even when individual predictions are noisy. Such statistical stabilisation effects highlight a key advantage of large-scale volunteered geographic data, where "Big Data" dynamics can mitigate local inaccuracies.

Spatial accuracy is equally important. For aerial detections, precise georeferencing is inherent in the data, as each pixel already corresponds to planimetric coordinates within the orthophoto mosaic. For street-level detections, however, a separate projection pipeline with monocular depth estimation and triangulation was required (Section 4.3.3). In total, more than 17.9 million detected objects were georeferenced, with median displacements of about 5–7.5 metres. These values show that the projection works reliably at a general level but also highlight its limitations.

Because the depth maps provide only relative rather than absolute distances, each object can be shifted slightly along the camera ray depending on the estimated depth and camera direction. The accuracy of Mapillary's own metadata adds further uncertainty: the GPS position and compass bearing of each image are user-generated and may differ by several metres from the real camera location or point in a slightly wrong direction. Krylov et al. (2019) compared Mapillary with professional street-level imagery such as Google Street View and found that, while object-detection accuracy improves with larger numbers of images, geolocation accuracy remains weaker because of higher noise in camera position and orientation. This is consistent with the observations here and underlines that spatial precision is difficult to achieve with crowdsourced imagery.

As a result, the projected features are generally in the correct direction but not always in the exact place. They are often located on the right street segment yet a few metres offset, or occasionally on the opposite side of a junction. This pattern is visible in the maps, where detections form plausible clusters but rarely align perfectly with the actual curb or crossing geometry. Such deviations are acceptable for analyses at the street-segment or district scale, where local errors tend to average out, but they limit applications that require precise object-level coordinates. Compared to aerial detections, which have sub-metre accuracy from the orthophotos, street-level projections therefore remain approximate. Even so, they considerably improve spatial alignment compared with using raw image coordinates and ensure that detected objects appear in the correct general position and orientation within the pedestrian network.

Such deviations are acceptable for analyses at the scale of street segments or neighbourhoods, where local positional errors tend to average out. However, they limit applications that require precise object coordinates, such as infrastructure inventories or automated asset management. In contrast, aerial detections inherit the sub-metre positional accuracy of the orthophotos, whereas street-level projections remain approximate. Even so, they still offer a clear improvement over raw image coordinates, ensuring that detected objects are placed in the correct general position and orientation within the pedestrian network.

This trade-off between directional correctness and metric precision is typical for monocular projection methods and has been discussed in previous research. Similar approaches have been proposed by Alzate et al. (2019), who combined object detection with geometry-based methods to project features onto maps. Their experiments using Mapillary data showed that detection accuracy improved with the number of available images, approaching that of professional street-level imagery such as Google Street View. However, the estimated object positions remained notably less accurate due to the higher noise in camera coordinates and orientation inherent to crowdsourced data (Krylov et al. 2019). Subsequent research by Chen et al. (2019) demonstrated that monocular depth estimation can be enhanced when semantic information is incorporated, as this leads to more coherent representations of object boundaries and urban structures. These studies confirm the general principle underlying the present approach: even if absolute positions

remain uncertain, depth-based projection substantially increases spatial coherence and provides sufficient positional reliability for aggregated, city-scale analyses.

Taken together, the findings show that reliable detection is possible for large and clearly visible features, while smaller or more subtle elements remain difficult. Aerial imagery allows very accurate mapping of structural features, street-level imagery adds important details at pedestrian scale, and crowdsourced data extends coverage with stable results when used in large quantities. The combination of these approaches provides a robust basis for further safety scoring and routing, even if trade-offs between accuracy, coverage, and detail remain. It should also be noted that not all safety-relevant infrastructure could be detected. Elements such as tactile paving for the visually impaired, narrow or poorly marked bicycle lanes, small stop lines, or missing and damaged signage were either not present in the datasets or could not be recognised reliably. This shows that the current framework can only capture part of the infrastructure relevant for school-route safety. In summary, RQ1 is answered by showing that large, distinct features can be detected reliably—particularly from aerial imagery—while smaller or subtle elements remain challenging in street-level data.

Main Takeaways

- Street-level models: good performance only for salient elements such as cars and traffic lights (AP50 $\sim$0.70); sidewalks and crossings medium (0.30–0.48); bike lanes and stop lines remain weak ($<$0.15).
- Aerial orthophotos: very high accuracy for structural features (mAP50 up to 0.98; precision/recall $\sim$0.99/0.95), confirming their suitability for network-level mapping.
- Crowdsourced data: the use of 1.2 million Mapillary images stabilised results, showing the Big Data effect via the Central Limit Theorem despite heterogeneous quality.
- Georeferencing: 17.9 million detections were shifted with median offsets of 5–7.5 m, allowing reliable integration of objects into the pedestrian network.
- Summary: Reliable detection is possible for large and distinct features, especially from aerial imagery, while smaller or occluded elements remain difficult to capture in street-level data.

6.2 Safety Scoring Approaches

Focus RQ2

To what extent can machine-learning and rule-based scoring approaches provide reliable and interpretable assessments of school-route safety, and what are their respective strengths and limitations?

Building on the established detection results and data foundation from Research Ques-

tion 1, the next step concerns how these extracted features can be translated into meaningful safety assessments. Research Question 2 therefore focuses on the development and evaluation of machine-learning and rule-based approaches for safety scoring.

The findings of this thesis indicate that both approaches are generally suitable for assessing the safety of school routes, yet their reliability is constrained. This reliability depends not only on technical performance but also on the quality of the input data, the clarity of categories, and the extent to which labels represent genuinely distinct safety conditions. As traffic safety research repeatedly shows, the strength of any classification system is tied to the strength of its underlying labels. Overlapping or subjective categories can substantially reduce accuracy (Silva et al. 2020). Under such conditions, false positives and false negatives are difficult to avoid, particularly when evidence in the imagery is ambiguous.

The machine learning (ML) approach illustrates both the promise and the pitfalls of data-driven models. Its main strength lies in producing probabilistic assessments: each segment receives not only a predicted class but also a probability distribution across safety levels. This avoids simplistic binary labels and explicitly communicates uncertainty. Rasterising YOLO detections into heatmaps before linking them to street segments reduced the influence of isolated false positives and better captured context (e.g., crossings, traffic lights). At the *city and district scale*, the aggregated ML scores reproduce plausible spatial patterns: peripheral districts tend to score higher than central hubs, and major nodes (e.g., Bellevue, Central) emerge as exposed structures in the ML maps (see Figures 5.17, 5.18 in the Chapter Results). These aggregates indicate that ML outputs are robust once spatially averaged, even if individual segments remain noisy.

At the same time, the limitations are non-trivial. The training data from the Zurich police were imbalanced and partly inconsistent. The labels were manually assigned by multiple people, introducing subjective variation in how borderline situations were classified. Even with shared guidelines, not all people interpret visual cues or contextual risks in exactly the same way, which adds noise to the ground truth and constrains model reliability. As Iranitalab et al. (2017) note, ML often outperforms traditional models, but predictive reliability depends heavily on representative class balance—especially in safety contexts where the rare cases matter most. Boundaries between categories (e.g., “safe” vs. “not recommended”) can be subjective and sometimes not visually discernible (Jain et al. 2020), setting an upper bound on achievable accuracy regardless of algorithmic tweaks. Finally, while probabilities and uncertainty intervals are valuable for experts, they are harder to explain to lay audiences (parents, schools).

The rule-based approach directly addresses interpretability. Every bonus/penalty is traceable to a visible feature, which aligns well with participatory planning and audit logic (Savolainen et al. 2011). Rasterisation again adds robustness by applying rules to local densities rather than isolated detections. In the *maps and statistics*, the rule-based scores tend to be more conservative near tram corridors, missing crossings, or busy

arterials, producing spatial patterns that broadly align with official assessments at the district scale (see Figures 5.17 and 5.18; Tables 5.11). This consistency at aggregated scales supports their use as an interpretable baseline for city-wide diagnostics.

These differences become particularly visible in the large-format overview maps (see Chapter 5.5.1), which visualise the full city-wide safety gradients. Both layers reveal a coherent city-wide gradient of perceived safety: outer residential districts such as 10, 11, and 12 appear predominantly blue, indicating higher scores, whereas central areas—particularly District 1 (Altstadt), District 4 (Langstrasse), and the main transport hubs around Central, Bellevue, and Escher Wyss—show extended orange and red corridors of lower safety. This overall pattern is consistent in both approaches, confirming that broad-scale structures are robust across modelling methods.

The **machine learning map** shows a finer and more textured surface, with localised areas of low scores along major crossings and tram corridors and smoother transitions within quieter quarters. It displays clearer safety gradients along main streets, where exposure is higher, but shows relatively little contrast between sidewalks and crossings. Strongly unsafe segments (deep red) are rare, and the map conveys more uncertainty in intermediate tones. This reflects how the ML scoring smooths local variation through probabilistic averaging: outliers are dampened, which increases overall stability but reduces local contrast. As a result, differences between crossings and adjacent sidewalks become less visible, illustrating the trade-off between sensitivity and robustness inherent in probabilistic models.

The **rule-based map**, by contrast, produces sharper, more categorical boundaries: entire street sections are either penalised or rewarded depending on rule thresholds, resulting in a discretised and visually segmented pattern. In this representation, most sidewalks are classified as generally safe, while crossings are systematically rated as risk points, especially where pedestrian priority or protection is lacking. This becomes evident in quieter quarters, where large portions of sidewalks appear blue, interrupted only by isolated red segments at crossings or intersections. As a result, the rule-based map appears less spatially uniform than the ML surface but is easier to interpret, since each colour can be directly linked to specific, rule-defined features. The approach thus provides higher transparency but a simpler and more rigid classification. While visually more categorical, the rule-based surface also tends to over-penalise crossings and understate variation along contiguous sidewalks. This behaviour is a direct outcome of fixed weight thresholds that ignore contextual interactions between features—such as traffic density or intersection complexity. Consequently, some segments appear uniformly safe simply because they lack penalised objects, even if the real-world environment would suggest residual risk.

For both approaches, class breaks were first calculated using the Jenks natural breaks algorithm (Jenks 1967) to match the distribution of safety scores. The breakpoints were then slightly adjusted so that the ML and rule-based maps used similar value ranges (around 0–20, 20–50, 50–60, 60–70, and 70–100). This ensures that visible differences

between the two maps come from the methods themselves rather than from unequal classification. A red–blue colour scheme from Harrower et al. (2003) was used for colour-blind accessibility and intuitive interpretation, with blue indicating safer areas and red showing less safe ones. Although this harmonisation facilitates visual comparison, it also introduces a cartographic bias: visual similarity may imply methodological agreement where underlying scores actually diverge. For example, two areas might both fall into the same "safe" class (70–100) and therefore appear equally blue on the map, even though their underlying safety scores differ substantially, such as 72 versus 98. In several central districts, absolute differences between ML and rule-based means exceed 30 points, indicating that the maps align in spatial pattern but not in magnitude.

A closer inspection of the central area (District 1) illustrates the main contrast between both scoring methods. The ML map shows greater local variation, identifying several safer side alleys within the old town and around the Botanical Garden. The rule-based map, by contrast, classifies most of the central network uniformly as "medium risk". Both representations successfully highlight key exposure areas while also revealing where input coverage is incomplete, such as semi-private or pedestrian-only passages that appear uncoloured due to missing data. In this sense, the maps not only visualise model outcomes but also help to diagnose data gaps and priorities for further collection.

The rule-based approach relies on fixed weights that cannot adapt to local context or learn from new evidence. Its deterministic scoring creates a false impression of precision and often compresses results into mid-range values, which can hide subtle variations or overlook local hotspots. These methodological differences also influence how each model translates safety scores into routing decisions.

A quantitative comparison between the SHAP-derived feature importance (Figure 5.15) and the manually defined weights (Tables 4.5 and 4.6) shows strong agreement for key traffic-related factors. In the SHAP ranking, the three most influential predictors are *dashed road markings* (mean |SHAP| $\approx$ 0.021), *traffic lights* ($\approx$ 0.018), and *tram infrastructure* ($\approx$ 0.014). These correspond closely to the highest rule-based penalties: "crossing with tram (no pedestrian protection)" (+6.0), "crossing with tram" (+3.0), and "crossing without marking" (+3.0). This alignment indicates that both approaches identify the same core sources of pedestrian risk: unprotected crossings, tram conflicts, and missing markings.

Beyond these overlaps, the SHAP analysis highlights several contextual variables that the rule-based model does not include. For example, *segment length* (mean |SHAP| $\approx$ 0.009), *vegetation/nature raster* ($\approx$ 0.004), and *motorised vehicles on road* ($\approx$ 0.004) all have a noticeable effect on predictions. These variables capture environmental context: longer segments and higher vehicle presence tend to increase predicted risk, whereas greener surroundings slightly reduce it.

Overall, both models rely on a similar structural core but differ in how they represent contextual variation. The rule-based system applies discrete thresholds and fixed penal-

ties, producing sharper contrasts but ignoring gradual transitions between conditions. The ML model, by contrast, reconstructs comparable patterns continuously, allowing subtle differences in exposure or urban form to modulate safety scores. These characteristics reflect their complementary strengths. The ML model adds probabilistic nuance and captures subtle spatial gradients, while the rule-based approach ensures transparent and reproducible results. Each has typical weaknesses: ML may smooth out local risks, whereas rule-based scoring can oversimplify context through fixed thresholds. Still, both consistently reproduce the main city-wide safety patterns, confirming the general validity of the framework. They also reveal meaningful spatial trends, such as identifying exposed intersections, hazardous corridors, and generally safer residential areas. Recent studies in road safety modelling have drawn similar conclusions, suggesting that combining interpretable rule-based frameworks with adaptive ML components can make use of the strengths of both approaches (Iranitalab et al. 2017; Zhang et al. 2018). In this context, the results of this thesis point in the same direction. This contrast reflects the conceptual distinction introduced in the Chapter Theoretical Framework, where rule-based approaches were described as transparent yet rigid and machine-learning approaches as adaptive but opaque, with interpretability methods such as SHAP providing a link between the two.

Main Takeaways

- Both methods provide reliable and interpretable safety assessments at the city scale, though local accuracy depends on data quality and feature coverage.
- ML captures spatial nuance and expresses uncertainty but can smooth over small-scale risks.
- Rule-based scoring is transparent and consistent but rigid and less sensitive to context.
- Combined use of both approaches offers the best balance between clarity and detail.

6.3 Routing and Practical Relevance

Focus RQ3

How do safety-based routing models perform in generating feasible and safer school-route alternatives compared to conventional shortest-path routes?

After establishing safety scores for each network segment, the third research question focuses on their practical use in routing. It investigates how integrating safety into pathfinding affects the resulting school routes.

The routing experiments demonstrate that safety-based models can produce viable alternatives to conventional shortest-path routing, although the magnitude of improve-

ment varies with local context. In the examples presented in Chapter 5, safer routes replaced exposed road segments with slightly longer alignments that prioritised signalised crossings or calmer side streets. These adjustments increased overall safety with only modest additional distance, making them realistic for everyday school use.

The trade-off between distance and safety follows a concave, non-linear pattern. Empirically, the results from all three case routes show that small detours of only one to three percent in distance produced disproportionately large safety gains of around four to seven percent. For example, in Route 1 (ML model), a +2.9% longer path yielded a +6.4% higher safety score, and in Route 2 (rule-based model), a +1.5% detour already increased safety by +4.0%. Beyond this point, longer detours of around five to six percent added little additional safety benefit. This means that most of the achievable safety gain occurs within a relatively short extra distance, while further improvements require much higher travel costs. The relationship is therefore clearly non-linear, reflecting a Pareto-like effect similar to the 80/20 principle, which states that most of the benefit can be achieved with relatively small detours, whereas the final increments require disproportionate effort (Men et al. 2022). A similar principle has been formulated by Hannah et al. (2018), who modeled pedestrian path safety as a weighted sum of distance and crash risk within a pavement network. Their model, grounded in multi-criteria decision-making theory, assumes that pedestrians behave rationally by minimising both length and risk simultaneously. This conceptual framework supports the present findings: the observed non-linear, concave trade-off between distance and safety mirrors the same principle of utility optimisation, where small increases in distance can yield disproportionately large reductions in risk.

The weighting parameters α (distance) and β (safety) control this balance. Low β values keep routes close to the fastest path, providing only marginal safety improvements, while moderate β values achieve the best compromise between safety and distance. The extreme case $\alpha = 0$, $\beta = 1$ optimises purely for safety and disregards distance entirely, leading to very long detours (e.g. the *Absolute Safest* variant of Route 3 in the ML network, Figure 5.25). Conversely, $\alpha = 1$, $\beta = 0$ reproduces the conventional shortest path. The most useful outcomes were found between these extremes, where balanced weights produced routes that were both safer and still practical.

Across the three case studies (Figures 5.23–5.25), the patterns were consistent. For Route 1 (Bahnhof Enge–Langstrasse), both methods achieved noticeable safety gains with modest detours, while the ML model showed a stronger sensitivity to crossings near *Helvetiaplatz*. Route 2 (Enge–Schwamendingen) covered a wider urban transect: here, the ML model avoided the dense *Bellevue* intersection altogether and selected a safer alignment via *Rathausbrücke*, whereas the rule-based network kept most variants along the Bellevue corridor except for the *Absolute Safest* route. Route 3 (Unterstrass) showed very similar paths in both models.

Safe crossings thus emerged as the decisive factor: if no connector with a safe crossing exists nearby, the algorithm cannot generate a safer route and instead reveals a network

gap. Where parallel corridors or additional crossings are available, however, the routing effectively shifts paths towards higher-scoring segments, highlighting the influence of local connectivity. For this reason, several preprocessing steps were taken to improve the underlying pedestrian network, as described in Section 5.1. These included merging disconnected segments, adding missing crossing links, and correcting network topology to ensure that safe alternatives were actually reachable. The improvements proved essential, as shown in the results, where routes in areas with enhanced crossing connectivity displayed more realistic and diverse alternatives compared to unrefined network sections. This also relates to route length and spatial context: where the local network is dense and well connected, the model can meaningfully explore and compare several alternatives; where connectivity is limited, only few or no safe options exist. Consequently, route length becomes another important determinant—short trips offer few alternatives, while longer ones provide greater flexibility and yield larger relative safety gains.

The comparison between the rule-based and ML-based routing confirms complementary strengths. The routing algorithm itself was identical in all tests; differences in the resulting paths are caused solely by the underlying input scores. Rule-based routing ensures clarity and reproducibility, while ML-based routing adds contextual sensitivity and captures subtle variations, though with more local variability. The balanced and safest variants of Route 2 illustrate this difference particularly well (see Figure 6.1 below). The ML network consistently avoided the exposed *Bellevue* crossing and selected the safer *Rathausbrücke*, whereas the rule-based network used the Bellevue corridor in both the *Balanced* and *Safest* variants, switching to Rathausbrücke only for the *Absolute Safest* route. In this example, the additional detour was modest—approximately +3.4% distance for the ML model and +1.5% for the rule-based model—while average safety improved by +7.1% and +4.0% respectively (see Tables 5.12 and 5.13). This shows that the method captures meaningful differences in safety perception at key decision points, even if the effects are moderate rather than dramatic.

As the Route 2 example illustrates, the most convincing results were those with small detours and a clear behavioural rationale—such as using a signalised crossing instead of an unsignalised one. At the city scale, the routing outputs also provide diagnostic insights by highlighting missing links, unsafe barriers, and areas where minor infrastructure interventions could unlock much safer paths. The main practical value therefore lies less in prescribing exact routes and more in identifying critical decision points and infrastructure gaps.

These insights also connect to the practical implementation of the routing tool. The routing module was built as a `FastAPI` web service and is accessible through a custom QGIS processing script. This setup lets planners compute and visualise safety-weighted routes directly within the GIS environment, ensuring consistent evaluation of multiple origin–destination pairs without manual work. The API returns GeoJSON outputs that can be styled and analysed automatically. Because the backend and data structures are

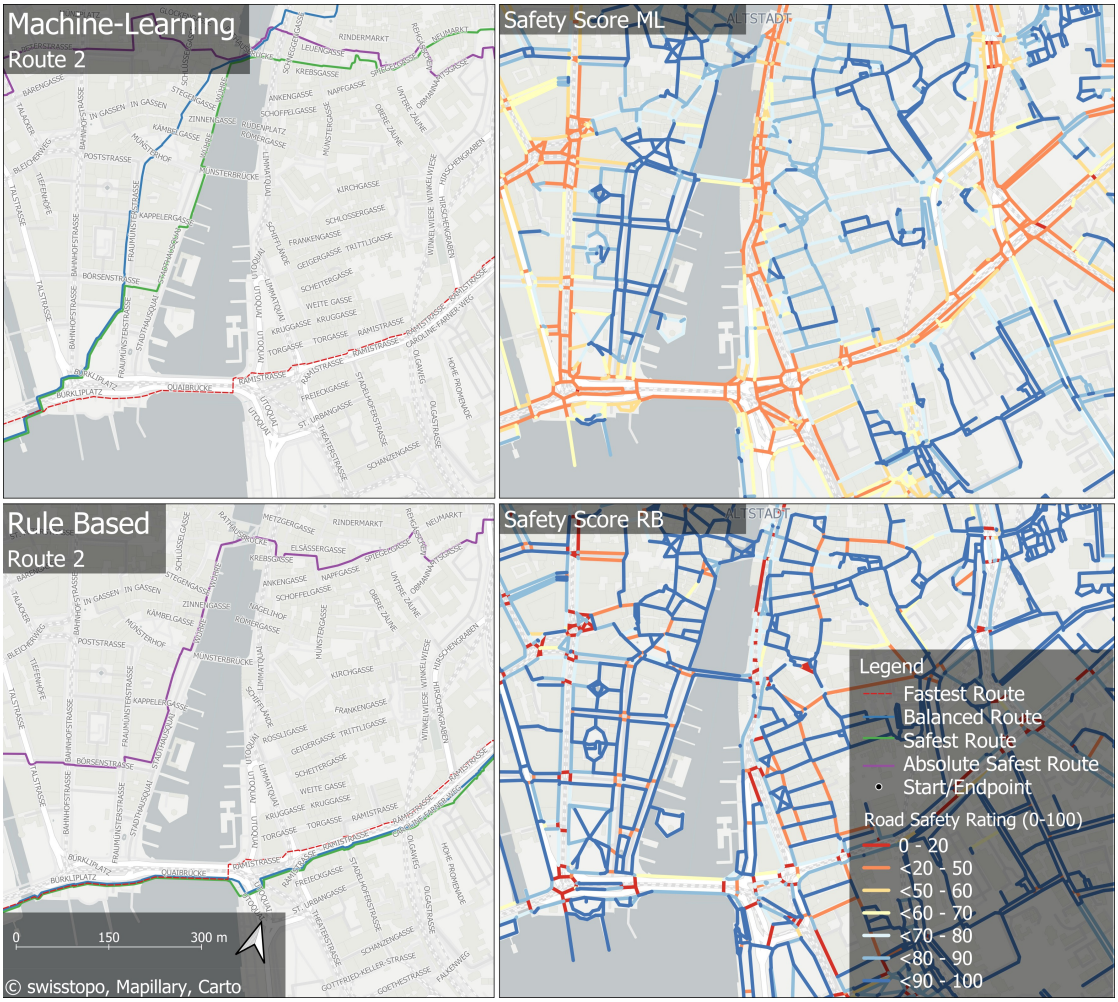


Figure 6.1: Zoomed-in view of Route 2: ML (top) vs. rule-based (bottom) with their respective safety-score maps.

modular, the same service could also run in WebGIS systems. This would enable public or participatory applications where users can explore safer route options interactively. Overall, the framework already provides the essential components for such extensions, including automated computation, API communication, and visualisation.

Main Takeaways

- Safety-based routing can generate safer alternatives with modest detours, particularly at major crossings and junctions.
- Most safety gains follow a Pareto-like pattern: small detours yield large improvements, while further gains require disproportionate costs.
- The routing algorithm was identical in all tests; differences between rule-based and ML results arise solely from their input scores.
- The API and QGIS integration enable automated, reproducible evaluation and could readily be extended to interactive WebGIS routing tools.
- In practice, the main value lies in highlighting critical decision points and infrastructure gaps rather than prescribing exact routes.

6.4 Comparison with Existing Plans and Stakeholder Feedback

Focus RQ4

How does the automated assessment compare to existing school route plans (e.g., Zurich's manually evaluated maps), and what insights emerge from stakeholder feedback?

The comparison with Zurich's official School Route Safety Plan reveals both alignments and divergences between automated and manual assessments. At the city and district scale, both approaches identify similar patterns: peripheral residential quarters are generally safer, while central hubs such as Bellevue, Central, and Escher Wyss consistently emerge as demanding or critical. This correspondence confirms that the automated framework captures the main structures of school route safety already recognised in practice. Differences appear at finer spatial scales. The ML-based method often assigns higher scores to calm side streets, reflecting its sensitivity to contextual features such as traffic lights or crossings. By contrast, the rule-based approach—closer in logic to the official plan—penalises tram corridors, missing crossings, or busy arterials more systematically, resulting in more conservative ratings. In several case examples, the automated framework also revealed potential hotspots not explicitly flagged in the official plan, such as underpasses or complex junctions. These divergences suggest that automation can add diagnostic value by surfacing blind spots or inconsistencies in manual evaluations.

Feedback from Fussverkehr Schweiz (Pascal Rengli et al. 2025) added a valuable practical perspective to the interpretation of the results. The technical implementation was acknowledged as innovative, while at the same time some methodological aspects appeared less convincing when looking at specific situations in Zurich North. For instance, the inclusion of private property connections, such as the courtyards of the Mötteliweg housing estate, raised the question of whether such paths should be considered at all,

given that they are only relevant for residents. At complex intersections such as Sternen Oerlikon, the automated assessment may underestimate the level of risk. This effect is partly linked to limited image coverage in the area, since few cars pass through and Mapillary imagery is concentrated along major traffic axes. In some traffic-calmed areas, by contrast, risks may appear somewhat overstated. Individual features, such as motorway feeder signage, were also noted to have a stronger influence on the scores than would be expected from the actual exposure of children. Finally, longer streets sometimes appeared rather homogeneous in the results, which could obscure local variations—for example on Baumackerstrasse, where perpendicular parking manoeuvres might pose higher risks than longitudinal parking.

Based on this feedback, the algorithm was subsequently refined to better recognise and evaluate pedestrian crossings and other safety-relevant features. The updated version improved local accuracy in several test areas, particularly in situations involving complex junctions or crossing points. A summary of these adjustments was shared with the stakeholder as part of the follow-up communication. Moreover, the City Police of Zurich expressed strong interest in the approach and will be presented with the refined results in a forthcoming meeting after the submission of this thesis.

Overall, this feedback illustrates both the potential and the limitations of the automated framework. While it is able to reproduce city-wide patterns consistently, local expertise remains indispensable for interpreting site-specific situations. Automated results should therefore be seen as a complement to participatory assessments: they provide systematic coverage and diagnostic insight, while stakeholder input ensures that contextual nuances and local knowledge are appropriately reflected. Together, these perspectives can contribute to more robust and widely accepted strategies for school route safety.

Main Takeaways

- The automated assessment shows general consistency with Zurich's official plan but also notable local deviations.
- It highlights additional or differently rated risk areas, reflecting its distinct data-driven logic.
- Stakeholder feedback emphasised both methodological innovation and the need for contextual interpretation, which informed subsequent algorithmic refinements.
- The City Police of Zurich expressed interest in the approach, underlining its practical relevance.
- Automated results should be integrated with expert review to ensure robustness and acceptance.

6.5 Research Aim

Focus Research Aim

The aim of this thesis is to design and evaluate an automated, computer vision–based framework to systematically assess the safety of school routes in the City of Zurich.

The overarching aim of this thesis was to develop and test an automated framework capable of assessing school-route safety in a consistent and data-driven manner across the entire urban area of Zurich. This final section brings together the findings from detection and data basis (Section 6.1), safety scoring approaches (Section 6.2), routing and practical relevance (Section 6.3), and comparison with existing plans and stakeholder feedback (Section 6.4) to evaluate how far this aim has been achieved, what the main contributions are, and where the framework still falls short.

Integration of the Research Questions

The following section summarises the main findings of each research question in relation to the overall aim.

The detection and data basis (Section 6.1) formed the technical foundation of the framework and represents both a central strength and a major limitation. It demonstrated that large-scale computer vision can successfully detect and map many safety-relevant features across the city, creating an empirical basis for automated safety assessment. At the same time, the overall precision of the street-level models remained modest, with mean average precision values around 0.25 for complex, heterogeneous imagery. In contrast, the aerial models achieved substantially higher accuracies (mAP50 up to 0.98), confirming that orthophotos are highly effective for mapping large, structural features such as markings and tram tracks. This combination of perspectives proved essential for capturing school-route safety at multiple spatial levels. Aerial data provided broad structural accuracy, while street-level imagery added detailed coverage at the pedestrian scale.

Nevertheless, smaller or visually subtle elements, such as worn markings or narrow bicycle lanes, were often missed, limiting the level of detail that can be represented in later stages. The processing of more than 1.2 million street-level images demonstrated, however, that large data volumes can stabilise results when aggregated: even imperfect detections produced meaningful safety patterns at the network and district level. This scalability shows that quantity and coverage can partly compensate for lower per-image precision, which is a valuable insight for future large-scale applications.

Building on this, the safety scoring approaches (Section 6.2) transformed these detections into interpretable measures of safety. The two scoring methods addressed different but complementary goals. The rule-based approach offered transparency and reproducibility, which are crucial for acceptance in planning contexts, while the machine-

learning model added flexibility and probabilistic nuance. Both approaches were, however, affected by the limitations of the underlying data and by subjective category boundaries in the training set. The imbalance between safe and unsafe examples in the Zurich police data biased the ML model toward generalizing predictions, while fixed thresholds in the rule-based system limited sensitivity to local context. Despite these weaknesses, the comparison of both approaches proved valuable by showing how different methodological choices influence spatial outcomes and by highlighting the trade-off between interpretability and adaptability.

The routing experiments (Section 6.3) demonstrated how these safety scores could be translated into practical, route-level applications. By balancing distance and safety, the models generated alternative paths that avoided hazardous segments with only minor additional distance. This worked particularly well at major crossings and junctions, where distinct safer options existed. At the same time, the experiments revealed several constraints. In areas with incomplete network connectivity or missing crossing links, the routing failed to suggest safer alternatives, thereby exposing gaps in data or infrastructure. In addition, the influence of the underlying safety map was sometimes uneven. Longer routes with few alternatives showed little change, whereas in dense areas the model occasionally generated overly cautious detours. These limitations underline that the usefulness of routing outputs depends primarily on the completeness and accuracy of the safety scores rather than on the algorithm itself.

The comparison with existing school route plans and stakeholder feedback (Section 6.4) provided an external perspective. The automated results broadly matched the spatial safety patterns identified by manual evaluations, confirming that the approach captures key structures of school route safety. At the same time, local discrepancies such as the inclusion of semi-private paths or the underestimation of complex intersections showed that contextual expertise remains essential.

Following the stakeholder feedback, the algorithm was refined to better identify and interpret pedestrian crossings and junctions, improving local accuracy in several test areas. A summary of these adjustments was shared with Fussverkehr Schweiz, who emphasised that while the framework's systematic coverage and diagnostic potential were recognised, interpretation must remain grounded in local knowledge. Overall, this feedback underlined that the greatest value of automation lies not in replacing expert judgement but in extending its reach through consistent, data-driven evidence.

The framework can therefore serve as an additional decision support layer, providing an initial quantitative assessment that helps experts and planners prioritise interventions and improve safety for children on their way to school.

Patterns and Scales of Analysis

Looking across all research questions, the framework performs differently depending on the spatial scale of analysis. At the city and district level, the aggregated results produced stable and interpretable patterns that aligned well with known traffic and

land-use characteristics (Figures 5.17 and 5.18; Table 5.11). Peripheral districts such as Höngg and Witikon appeared generally safer, reflecting calmer street networks and fewer complex intersections, whereas central areas such as Altstadt, Oerlikon, and Escher Wyss showed higher exposure and more fragmented safety structures. These contrasts indicate that the automated results capture meaningful large-scale variations rather than random artefacts. At this strategic level, the framework provides a useful diagnostic tool for identifying districts where children face higher risks and where targeted interventions could have the greatest effect.

At finer spatial scales, the strengths and weaknesses of the framework become more apparent. It successfully distinguished generally safer from more exposed intersections and corridors, indicating where infrastructure adjustments or alternative connections could be most effective. However, it is less sensitive to very small-scale variations within the same location. For example, while the model can reliably identify that a crossing is safer than a nearby unsignalised junction, it cannot yet capture whether a pedestrian crossing a few metres further ahead or behind would experience a different level of safety. This limitation arises from the detection and aggregation process, which generalises conditions along continuous street segments rather than modelling micro-level spatial detail. As a result, the framework captures overall safety patterns and hotspot locations but may overlook subtle positional or behavioural nuances. Therefore, while it performs well for aggregated analysis and intersection-level diagnostics, it should not yet be used for precise, edge-by-edge risk evaluation or as a stand-alone planning instrument. Its current role is best understood as a diagnostic layer that can inform, but not dictate, planning decisions.

Reflections on the Aim

Taken together, the findings indicate that the overarching research aim has been achieved in conceptual and methodological terms, but only partly in empirical performance. The thesis demonstrates that an automated, computer vision–based framework can be built, scaled, and meaningfully applied to urban school-route safety. It connects detection, scoring, and routing into a coherent workflow that produces interpretable outputs across multiple spatial levels. This represents a significant contribution in terms of technical integration and methodological design. However, the accuracy and practical validity of the results remain limited by data quality, class imbalance, and interpretative uncertainty. In that sense, the framework should be regarded as a functional prototype rather than a finished assessment tool. Quantitatively, the framework achieved moderate empirical reliability, with mean detection accuracies around 0.25 mAP50 and typical safety gains of 4–7 % in route comparisons, demonstrating proof of concept but leaving room for refinement.

The empirical results mirror broader challenges noted in urban computer vision. Naik et al. (2017) show how automated image-based models can successfully quantify urban change, yet also emphasize that such methods simplify visual complexity and cannot

capture the full social and contextual meaning of places. Similarly, Ding et al. (2021) underline that even large-scale, well-annotated datasets for aerial object detection remain limited by orientation, scale variation, and nonuniform densities, constraining precision in real-world applications. These parallels confirm that the modest empirical performance of the Zurich framework reflects structural challenges in the field rather than implementation flaws: the limitations stem from the inherent complexity of urban scenes and the representativeness of the available imagery, not merely from algorithmic shortcomings.

At the same time, the research highlights a clear tension between automation and interpretability. As Kitchin (2017) points out, algorithmic systems in urban analysis can provide consistent and large-scale results, but they also tend to hide the assumptions and data biases behind them. The same applies here: the automated outputs are systematic and reproducible, yet they sometimes miss the local context that human observers would easily recognise. The transparent logic of the rule-based model helps to reduce this problem, and the integration of SHAP-based feature attribution further increases interpretability on the ML side. This allows individual model decisions to be traced back to specific input features—revealing, for example, how crossings, tram infrastructure, or markings contribute to predicted safety scores. Consequently, the ML component no longer functions as a full black box but as a semi-transparent model whose logic can be partially inspected and validated. As Breiman (2001) already noted, ensemble methods such as Random Forests improve predictive performance at the cost of interpretability, but techniques like SHAP now help to bridge this gap by making the underlying decision structure more explainable.

Viewed in a planning context, the research aligns with broader debates on how quantitative automation and participatory interpretation can complement each other. Goodspeed (2020) highlights that planning for uncertain urban futures requires iterative, collaborative processes that integrate data-driven insights with stakeholder learning rather than replace it. In this sense, the present framework can be seen as a diagnostic tool supporting such iterative processes: it provides an analytical basis for dialogue, but not prescriptive solutions.

Main Takeaways

- The automated framework for assessing school-route safety was successfully developed and tested in Zurich, but its reliability remains constrained by data quality and class imbalance.
- Aggregated results at district and city level are stable and plausible, yet local accuracy varies with input completeness and detection precision.
- Rule-based and ML scoring complement each other; SHAP makes the ML component easier to interpret, though full transparency is still limited.
- Routing experiments showed practical potential for safer path suggestions.
- The framework offers a consistent, city-wide diagnostic layer, but it should be used as a decision aid rather than a stand-alone planning tool.

6.6 Limitations

While the results demonstrate the potential of automated safety assessment, the framework developed in this thesis is subject to several substantial limitations. These concern (i) the **data basis** and its representativeness, (ii) the **technical and methodological setup**, and (iii) the **conceptual boundaries** of what can be captured as "school-route safety." Figure 6.2 summarises these domains as overlapping spheres, each containing distinct yet interdependent constraints.

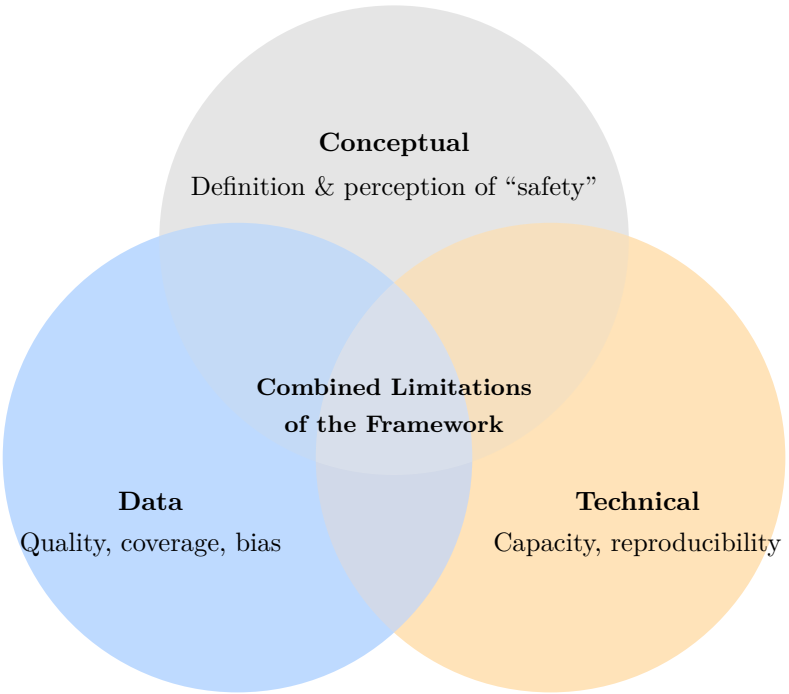


Figure 6.2: Overlapping spheres of limitation in the automated school-route framework (own figure).

6.6.1 Data Quality, Coverage, and Bias

The validity of the framework depends fundamentally on the quality and representativeness of its input data. As outlined in Section 6.1, the analysis uses crowdsourced street-level imagery from Mapillary, which is unevenly distributed across the city. Coverage is high along arterial roads but substantially lower in residential streets, precisely where children's mobility is most relevant. This imbalance reflects known biases in volunteered geographic information, where contributors tend to capture car-oriented or visually prominent environments (Thebault-Spieker et al. 2018; Quattrone et al. 2015). Because most Mapillary sequences are recorded from vehicles, pedestrian areas are often missing, which limits the capacity to detect child-relevant features such as narrow sidewalks or refuge islands.

Temporal inconsistencies further reduce reliability. Some imagery is outdated (Figure 3.3b) or affected by seasonal differences in light and vegetation, while temporary elements such as parked cars or construction barriers distort visibility . Aerial orthophotos provide complete coverage and positional accuracy but cannot represent dynamic or small-scale safety cues such as signal lights, markings, or signage.

Detection accuracy is also constrained by visual complexity. Even advanced models such as YOLO11 achieved only moderate precision for small or partly occluded objects, with an overall mAP50 of 0.238 (Table 5.2 and 5.3). This is partly due to limited training data and hardware capacity (see Section 6.6.2), but also to the inherent heterogeneity of urban imagery. Hence, results should be interpreted as indicative rather than exact.

Spatial projection accuracy remains another limitation. Street-level detections were georeferenced via monocular depth estimation and triangulation, yielding median positional offsets of roughly 5–7.5 m. While sufficient for district-level patterns, this margin is too large for intersection-scale diagnostics where even minor shifts can alter the safety classification of crossings or traffic islands. Such uncertainties propagate into the subsequent scoring and routing analyses but could not be formally quantified.

6.6.2 Technical and Methodological Constraints

The second limitation concerns computational capacity and methodological reproducibility. Deep-learning inference and training are resource-intensive; all experiments were executed on consumer-grade GPUs, constraining input resolution, batch size, and training depth. Processing over 1.2 million images required several hundred GPU hours, preventing extensive hyperparameter tuning or model comparisons. Hardware failures, memory overflows, and dependency conflicts further limited experimental stability. While an Anaconda-based Python environment offered flexibility, it lacks full reproducibility compared to containerised setups (e.g., Docker) (Boettiger 2015). A future implementation on scalable cloud infrastructure or distributed GPU clusters could substantially improve robustness and transparency (Zhou et al. 2024).

From a methodological perspective, both the rule-based and machine learning ap-

proaches were influenced by data imbalance and incomplete feature coverage. The police training dataset contained considerably fewer unsafe than safe examples, which may have biased the ML model towards optimistic classifications. The rule-based scoring, on the other hand, followed a consistent but somewhat rigid logic, occasionally overemphasising crossings or overlooking local context. Uncertainty propagation between detection, scoring, and routing was not yet modelled explicitly, which could be addressed in future iterations to improve interpretability. Reproducibility is also partly constrained because the computer vision models rely on third-party pretrained weights that may evolve over time.

Routing performance directly reflects these dependencies. Although the algorithmic implementation proved reliable, missing or uncertain safety scores sometimes restricted its ability to suggest alternative paths. Therefore, the routing results should be interpreted primarily as diagnostic outputs that highlight areas with lower data reliability or ambiguous safety patterns, rather than as prescriptive recommendations for navigation.

6.6.3 Conceptual Boundaries and the Meaning of "Safety"

A final limitation concerns the conceptual scope of what is represented as "school route safety." As introduced in Section 2.1, safety is not a fixed or purely technical condition but a socially and psychologically constructed perception. A route that appears objectively safe may still feel unsafe to a child because of noise, isolation, or unfamiliar surroundings, whereas a physically risky situation may feel safe when it is familiar or playful. This distinction between *measured risk* and *felt safety* highlights that the framework can approximate, but not fully represent, children's lived experience of safety (Iqbal 2023).

Developmental research shows that children's understanding of danger evolves gradually through observation, practice, and adult guidance rather than through infrastructure alone. Thomson et al. (2005) demonstrated that safe pedestrian behaviour depends on metacognitive skills such as anticipation, timing, and decision making, which mature with training and feedback. Barton et al. (2006) further showed that even brief, skill-based instruction can improve children's behaviour, but only when adults actively model and discuss safety decisions. Without such interaction, children rarely internalise safety cues, underscoring that safety is learned socially rather than embedded in the built environment.

Risk perception also varies across individuals. Morrongiello et al. (2007) emphasise that children's risk taking is shaped by cognitive appraisals, emotional motivations, and gendered socialisation, as boys often associate risk with excitement and competence, while girls are encouraged toward caution. Consequently, what feels safe for one child may feel unsafe for another, depending on age, cognitive maturity, temperament, and context (Barton et al. 2006; Morrongiello et al. 2007; Thomson et al. 2005). Parents further shape this perception: they tend to lead crossings silently, leaving few opportunities for children to practise decision-making themselves (Barton et al. 2006). Safety

is thus not an environmental property but a co-produced experience between physical setting, social modelling, and children's agency.

From an analytical standpoint, the framework operationalises safety as an infrastructural *potential for protection*, based on visible features such as sidewalks or crossings, but it omits the emotional, behavioural, and social layers emphasised in inclusive space research (Iqbal 2023). Automated scoring therefore reflects an adult, planner-centred notion of safety rather than the child's own sense of security. Recognising these conceptual boundaries is essential: the model quantifies exposure and protection but cannot capture fear, confidence, trust, or companionship—core dimensions of how safety is actually experienced.

Main Takeaways

- Uneven and vehicle-biased crowdsourced imagery causes spatial and temporal gaps, limiting detection reliability.
- Detection, scoring, and routing accuracy depend strongly on model design, data balance, and visual complexity.
- Limited computational capacity restricted model scale, hyperparameter tuning, and full reproducibility.
- Uncertainty propagation across detection, scoring, and routing remains unresolved.
- "School-route safety" is subjective and socially learned; automated metrics capture infrastructural risk but not children's lived sense of safety.

7 Conclusion

This thesis set out to assess school-route safety in the city of Zurich through a computer-vision-based analytical framework. The approach combined automated detection of safety-relevant features, spatial integration, and interpretable scoring to provide a systematic and city-wide assessment of the built environment surrounding children's school routes. By linking visual evidence with geographic data, the study explored how infrastructural and spatial conditions shape the everyday safety of walking children.

The results demonstrated that such an automated framework can approximate known spatial safety patterns and reveal previously unseen local variations. Both the rule-based and the machine-learning scoring approaches proved applicable, yet each with distinct advantages. The rule-based method ensured transparent and reproducible evaluation, while the machine-learning model captured gradual spatial transitions and contextual nuances. Their combination strengthened the interpretability and practical relevance of the findings. Aggregated results at the district level aligned with established planning knowledge, but local reliability remained constrained by image coverage, detection accuracy, and the heterogeneity of the urban fabric.

In addition, routing experiments illustrated how safety assessments can inform everyday mobility planning. Routes optimised for low risk often followed longer but calmer paths, demonstrating the trade-off between safety and efficiency. This highlights the potential of integrating safety scores into school mobility planning, while also underlining that such results require expert interpretation and should complement, not replace, on-site assessment.

Despite technical and data-related limitations, the study achieved its main goal: to translate the multidimensional concept of safety into a reproducible, data-driven framework. The approach demonstrates how computer vision can support planning practice by creating consistent, scalable, and transparent safety indicators across the entire urban network. It also contributes methodologically by combining automation with interpretability and by situating algorithmic analysis within a critical understanding of safety as a social and perceptual construct.

The framework developed here can serve as a diagnostic foundation for future studies and municipal applications. It can be extended to other Swiss cities or adapted to include dynamic data sources such as seasonal imagery or mobility traces. Ultimately, the study shows that computational tools, when used reflexively, can enhance both the analytical depth and the democratic transparency of urban safety assessment. They provide not a replacement for human judgement, but a new lens through which safer, fairer school routes can be envisioned and built.

7.1 Future Work

Building on the framework developed in this thesis, several directions for future work can be identified. A first step would be to apply the approach to other Swiss cities to test its adaptability under different urban conditions and data environments. Since the workflow is modular and open, it can be transferred to various spatial contexts with limited adjustments, enabling comparative analyses of school-route safety at regional or national scale.

Further development should also focus on enriching the data basis. Integrating municipal GIS layers such as traffic volumes, speed limits or accident records would allow for a more nuanced representation of exposure and risk. Including additional spatial or temporal attributes, for example through seasonal imagery or updated street-level data, could improve both coverage and timeliness.

Methodologically, the framework could be extended to better account for uncertainty and sensitivity in its outputs. Probabilistic techniques such as Monte Carlo simulations or Bayesian updating could be used to quantify confidence ranges in safety scores and routing results. This would enhance the interpretability of outcomes and support more robust decision-making in planning practice.

A further direction lies in making the framework more interactive and participatory. Developing a Web-GIS platform would make the analyses accessible to planners, schools and the public, while enabling shared exploration of safety patterns and routing options. Feedback collected through such participatory use could be systematically integrated into the model, allowing local knowledge and stakeholder perspectives to refine and update the assessment over time.

Statement of Authorship

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Digital tools were used to support parts of the writing and analysis process. Artificial intelligence tools, including OpenAI's ChatGPT, were used for language editing, stylistic revision, and occasional code refinement, and Grammarly was used for grammar checking and minor linguistic improvements. All conceptual, analytical, and interpretative content, including the research design, data analysis, results, and discussion, was developed independently by the author.

Zurich, October 22, 2025 **Claude Widmer**

Bibliography

Abdullah, Muhammad et al. (2023). "Signal-Free Corridor Development and Their Impact on Pedestrians: Insights from Expert and Public Surveys". In: *Sustainability*. DOI: 10.3390/su151914480.

Aio-Libs Community (2025). *Aiohttp: Asynchronous HTTP Client/Server for Asyncio and Python*.

Akyon, Fatih Cagatay, Sinan Onur Altinuc, and Alptekin Temizel (2022). "Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection". In: *2022 IEEE International Conference on Image Processing (ICIP)*, pp. 966–970. DOI: 10.1109/ICIP46576.2022.9897990.

Akyon, Fatih Cagatay et al. (Nov. 2021). *SAHI: A Lightweight Vision Library for Performing Large Scale Object Detection and Instance Segmentation*. Zenodo. DOI: 10.5281/zenodo.5718950.

Ali, Mohammad, Michael Emch, and Jean-Paul Donnay (June 2002). "Spatial Filtering Using a Raster Geographic Information System: Methods for Scaling Health and Environmental Data". In: *Health & Place* 8.2, pp. 85–92. ISSN: 13538292. DOI: 10.1016/S1353-8292(01)00029-6.

Alzate, Carlos et al., eds. (2019). *ECML PKDD 2018 Workshops: Nemesis 2018, UrbReas 2018, SoGood 2018, IWAISe 2018, and Green Data Mining 2018, Dublin, Ireland, September 10-14, 2018, Proceedings*. Vol. 11329. Lecture Notes in Computer Science. Cham: Springer International Publishing. ISBN: 978-3-030-13452-5 978-3-030-13453-2. DOI: 10.1007/978-3-030-13453-2.

Amt für Raumentwicklung (2025). *Linien Des Öffentlichen Verkehrs*. Open Data Portal Stadt Zürich.

Antwi, Richard et al. (June 13, 2024). "Turning Features Detection from Aerial Images: Model Development and Application on Florida's Public Roadways". In: *Smart Cities* 7.3, pp. 1414–1440. ISSN: 2624-6511. DOI: 10.3390/smartcities7030059.

Apache Arrow Developers (2025). *Apache Arrow: PyArrow*.

Bansal, Raghav, Gaurav Raj, and Tanupriya Choudhury (2016). "Blur Image Detection Using Laplacian Operator and Open-CV". In: *2016 International Conference System Modeling & Advancement in Research Trends (SMART)*, pp. 63–67. DOI: 10.1109/SYSMART.2016.7894491.

Barton, B. K., D. C. Schwebel, and B. A. Morrongiello (Oct. 3, 2006). "Brief Report: Increasing Children's Safe Pedestrian Behaviors through Simple Skills Training". In: *Journal of Pediatric Psychology* 32.4, pp. 475–480. ISSN: 0146-8693, 1465-735X. DOI: 10.1093/jpepsy/jsl028.

Bartzokas-Tsiompras, Alexandros and Efthimios Bakogiannis (Dec. 31, 2023). "Quantifying and Visualizing the 15-Minute Walkable City Concept across Europe: A Multi-

criteria Approach". In: *Journal of Maps* 19.1, p. 2141143. DOI: 10.1080/17445647.2022.2141143.

Beddow, Christopher and Contributors (2021). *Mapillary Python SDK – Python Library for Mapillary API V4*.

Biernat, Elżbieta, Justyna Krzepota, and Dorota Sadowska (2020). "Cycling to Work: Business People, Encourage More Physical Activity in Your Employees!" In: *Work (reading, Mass.)* DOI: 10.3233/wor-203091.

Boettiger, Carl (Jan. 20, 2015). "An Introduction to Docker for Reproducible Research". In: *ACM SIGOPS Operating Systems Review* 49.1, pp. 71–79. ISSN: 0163-5980. DOI: 10.1145/2723872.2723882.

Bradski, G. (2000). "The OpenCV Library". In: *Dr. Dobb's Journal of Software Tools*.

Breiman, Leo (Oct. 2001). "Random Forests". In: *Machine Learning* 45.1, pp. 5–32. ISSN: 0885-6125, 1573-0565. DOI: 10.1023/A:1010933404324.

Brodersen, Kay Henning et al. (Aug. 2010). "The Balanced Accuracy and Its Posterior Distribution". In: *2010 20th International Conference on Pattern Recognition*. 2010 20th International Conference on Pattern Recognition (ICPR). Istanbul, Turkey: IEEE, pp. 3121–3124. ISBN: 978-1-4244-7542-1. DOI: 10.1109/ICPR.2010.764.

Bundesverfassung der Schweizerischen Eidgenossenschaft (2021). *Bundesverfassung Der Schweizerischen Eidgenossenschaft Vom 18. April 1999 (Stand 7. März 2021)*. URL: https://fedlex.data.admin.ch/eli/cc/1999/404 (visited on 10/05/2025).

Burrough, P. and Rachael McDonnell (Jan. 1998). "Principle of Geographic Information Systems". In: ISSN: 1567-5777.

Carlson, Jordan et al. (2014). "Built Environment Characteristics and Parent Active Transportation Are Associated with Active Travel to School in Youth Age 12–15". In: *British Journal of Sports Medicine*. DOI: 10.1136/bjsports-2013-093101.

Chandra, Rakesh Vidya and Bala Subrahmanyam Varanasi (2015). *Python Requests Essentials*. Packt Publishing Ltd.

Chawla, N. V. et al. (June 1, 2002). "SMOTE: Synthetic Minority over-Sampling Technique". In: *Journal of Artificial Intelligence Research* 16, pp. 321–357. ISSN: 1076-9757. DOI: 10.1613/jair.953.

Chen, Po-Yi et al. (June 2019). "Towards Scene Understanding: Unsupervised Monocular Depth Estimation with Semantic-Aware Representation". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, pp. 2619–2627. ISBN: 978-1-7281-3293-8. DOI: 10.1109/CVPR.2019.00273.

Cieśla, M. and Elżbieta Macioszek (2022). "The Perspective Projects Promoting Sustainable Mobility by Active Travel to School on the Example of the Southern Poland Region". In: *Sustainability*. DOI: 10.3390/su14169962.

Clark, Alex (2015). *Pillow (PIL Fork) Documentation*. readthedocs.

Colvin, Samuel et al. (2025). *Pydantic Validation*. Version v2.12.0a1+dev.

D'Haese, Sara et al. (2011). "Criterion Distances and Environmental Correlates of Active Commuting to School in Children". In: *International Journal of Behavioral Nutrition and Physical Activity*. DOI: 10.1186/1479-5868-8-88.

Da Costa-Luis, Casper (May 2019). "Tqdm: A Fast, Extensible Progress Meter for Python and CLI". In: *Journal of Open Source Software* 4, p. 1277. DOI: 10.21105/joss.01277.

Davies, Tilman M., Jonathan C. Marshall, and Martin L. Hazelton (2017). *Tutorial on Kernel Estimation of Continuous Spatial and Spatiotemporal Relative Risk with Accompanying Instruction in R*. Version 1. DOI: 10.48550/ARXIV.1707.06888. URL: https://arxiv.org/abs/1707.06888 (visited on 10/10/2025). Pre-published.

Desprez, Marine, Kyle Zawada, and Daniel Ramp (Mar. 2022). "Overcoming the Ordinal Imbalanced Data Problem by Combining Data Processing and Stacked Generalizations". In: *Machine Learning with Applications* 7, p. 100241. ISSN: 26668270. DOI: 10.1016/j.mlwa.2021.100241.

Ding, Jian et al. (2021). "Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges". Version 2. In: DOI: 10.48550/ARXIV.2102.12219.

Dirks, Kim N., Jennifer Salmond, and Nicholas Talbot (2018). "Air Pollution Exposure in Walking School Bus Routes: A New Zealand Case Study". In: *International Journal of Environmental Research and Public Health*. DOI: 10.3390/ijerph15122802.

Dyck, Delfien Van et al. (2010). "Criterion Distances and Correlates of Active Transportation to School in Belgian Older Adolescents". In: *International Journal of Behavioral Nutrition and Physical Activity*. DOI: 10.1186/1479-5868-7-87.

Encode OSS (2025). *Uvicorn: ASGI Web Server Implementation*.

Federal Office of Topography swisstopo (2023a). *SwissBoundaries3D - Administrative Boundaries of Switzerland*.

— (2023b). *SWISSIMAGE 10 Cm - High-Resolution Orthophotos of Switzerland*.

— (2025). *swissTLM3D*.

Feudjio Tezong, Steffel Ludivin et al. (2024). "Investigating and Improving Pedestrian Safety in an Urban Environment of a Low- or Middle-Income Country: A Case Study of Yaoundé, Cameroon". In: *Future Transportation*. DOI: 10.3390/futuretransp4020026.

Fussverkehr Schweiz (Jan. 16, 2025). *Über Uns - Fussverkehr Schweiz*. Über Uns. URL: https://fussverkehr.ch/ueber-uns/ (visited on 10/05/2025).

Geofabrik GmbH (2025). *OpenStreetMap*. Geofabrik GmbH.

Gillies, Sean et al. (2013). *Rasterio: Geospatial Raster I/O for Python Programmers*. Mapbox.

Gillies, Sean et al. (2025). *Shapely*. Version 2.1.1. DOI: 10.5281/zenodo.5597138.

Goodspeed, Robert (2020). *Scenario Planning for Cities and Regions: Managing and Envisioning Uncertain Futures*. Cambridge, Massachusetts: Lincoln Institute of Land Policy. 240 pp. ISBN: 978-1-55844-400-3.

Gordon, Ariel et al. (2019). "Depth from Videos in the Wild: Unsupervised Monocular Depth Learning from Unknown Cameras". Version 1. In: DOI: 10.48550/ARXIV.1904.04998.

Gorriz, Juan M et al. (2024). *Is K-fold Cross Validation the Best Model Selection Method for Machine Learning?* Version 2. DOI: 10.48550/ARXIV.2401.16407. URL: https://arxiv.org/abs/2401.16407 (visited on 10/11/2025). Pre-published.

GRASS Development Team (2024). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.4*. manual. USA: Open Source Geospatial Foundation. DOI: 10.5281/zenodo.5176030.

Greer, Anna E. et al. (2019). "Walking toward a Brighter Future: A Participatory Research Process to Advocate for Improved Walk-to-School Corridors". In: *Health Promotion Practice*. DOI: 10.1177/1524839919890872.

Guryanov, Aleksei (Dec. 2019). "Histogram-Based Algorithm for Building Gradient Boosting Ensembles of Piecewise Linear Decision Trees". In: pp. 39–50. ISBN: 978-3-030-37333-7. DOI: 10.1007/978-3-030-37334-4_4.

Hagberg, Aric, Pieter Swart, and Daniel S Chult (2008). *Exploring Network Structure, Dynamics, and Function Using NetworkX*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Hannah, Charlotte, Irena Spasić, and Padraig Corcoran (Dec. 2018). "A Computational Model of Pedestrian Road Safety: The Long Way Round Is the Safe Way Home". In: *Accident Analysis and Prevention* 121, pp. 347–357. ISSN: 00014575. DOI: 10.1016/j.aap.2018.06.004.

Harris, Charles R. et al. (Sept. 2020). "Array Programming with NumPy". In: *Nature* 585.7825, pp. 357–362. DOI: 10.1038/s41586-020-2649-2.

Harrower, Mark and Cynthia A. Brewer (June 2003). "ColorBrewer.Org: An Online Tool for Selecting Colour Schemes for Maps". In: *Cartographic Journal* 40.1, pp. 27–37. ISSN: 0008-7041, 1743-2774. DOI: 10.1179/000870403235002042.

Herrador-Colmenero, Manuel, Emilio Villa-González, and Palma Chillón (2017). "Children Who Commute to School Unaccompanied Have Greater Autonomy and Perceptions of Safety". In: *Acta Paediatrica* 106.12, pp. 2042–2047. ISSN: 1651-2227. DOI: 10.1111/apa.14047.

Hoebel, Katharina V. et al. (2023). *A Generalized Framework to Predict Continuous Scores from Medical Ordinal Labels*. Version 1. DOI: 10.48550/ARXIV.2305.19097. URL: https://arxiv.org/abs/2305.19097 (visited on 10/11/2025). Pre-published.

Hugging Face (Oct. 2, 2025). *Hugging Face – the AI Community Building the Future*. URL: https://huggingface.co/ (visited on 10/05/2025).

Hunter, J. D. (2007). "Matplotlib: A 2D Graphics Environment". In: *Computing in Science & Engineering* 9.3, pp. 90–95. DOI: 10.1109/MCSE.2007.55.

Iqbal, Asifa (Dec. 20, 2023). "Inclusive, Safe and Resilient Public Spaces: Gateway to Sustainable Cities?" In: *Urban Transition - Perspectives on Urban Systems and*

Environments. Ed. by Marita Wallhagen and Mathias Cehlin. IntechOpen. ISBN: 978-1-83962-412-4 978-1-83962-413-1. DOI: 10.5772/intechopen.97353.

Iranitalab, Amirfarrokh and Aemal Khattak (Nov. 2017). "Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction". In: *Accident Analysis and Prevention* 108, pp. 27–36. ISSN: 00014575. DOI: 10.1016/j.aap.2017.08.008.

Islam, Mohammad (Feb. 2018). "Sample Size and Its Role in Central Limit Theorem (CLT)". In: 4, pp. 1–7.

Jain, Abhinav et al. (Aug. 23, 2020). "Overview and Importance of Data Quality for Machine Learning Tasks". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Virtual Event CA USA: ACM, pp. 3561–3562. ISBN: 978-1-4503-7998-4. DOI: 10.1145/3394486.3406477.

Jenks, George F. (1967). "The Data Model Concept in Statistical Mapping". In.

Jiang, Peiyuan et al. (2022). "A Review of Yolo Algorithm Developments". In: *Procedia Computer Science* 199, pp. 1066–1073. ISSN: 18770509. DOI: 10.1016/j.procs.2022.01.135.

Joblib Development Team (2020). *Joblib: Running Python Functions as Pipeline Jobs*.

Jocher, Glenn and Jing Qiu (2024). *Ultralytics YOLO11*. Version 11.0.0.

Kaya, Ömer, Muhammed Yasin Çodur, and Enea Mustafaraj (Apr. 18, 2023). "Automatic Detection of Pedestrian Crosswalk with Faster R-CNN and YOLOv7". In: *Buildings* 13.4, p. 1070. ISSN: 2075-5309. DOI: 10.3390/buildings13041070.

Khanam, Rahima and Muhammad Hussain (Oct. 23, 2024). *YOLOv11: An Overview of the Key Architectural Enhancements*. DOI: 10.48550/arXiv.2410.17725. arXiv: 2410.17725 [cs]. URL: http://arxiv.org/abs/2410.17725 (visited on 09/06/2025). Pre-published.

Khavarian-Garmsir, Amir Reza, Ayyoob Sharifi, and Ali Sadeghi (Jan. 1, 2023). "The 15-Minute City: Urban Planning and Design Efforts toward Creating Sustainable Neighborhoods". In: *Cities* 132, p. 104101. ISSN: 0264-2751. DOI: 10.1016/j.cities.2022.104101.

Kim, Young-Jae and Chanam Lee (Jan. 2020). "Built and Natural Environmental Correlates of Parental Safety Concerns for Children's Active Travel to School". In: *International Journal of Environmental Research and Public Health* 17.2, p. 517. ISSN: 1660-4601. DOI: 10.3390/ijerph17020517.

Kitchin, Rob (Jan. 2, 2017). "Thinking Critically about and Researching Algorithms". In: *Information, Communication & Society* 20.1, pp. 14–29. ISSN: 1369-118X, 1468-4462. DOI: 10.1080/1369118X.2016.1154087.

Klopper, Abigail (Oct. 3, 2024). "How Walkable Is Your City? Online Tool Shows How Major Centres Measure Up". In: *Nature* 634.8032, pp. 38–38. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/d41586-024-03145-3.

Krylov, Vladimir A. and Rozenn Dahyot (2019). "Object Geolocation from Crowdsourced Street Level Imagery". In: *ECML PKDD 2018 Workshops*. Ed. by Carlos Alzate et al. Vol. 11329. Cham: Springer International Publishing, pp. 79–83. ISBN: 978-3-030-13452-5 978-3-030-13453-2. DOI: 10.1007/978-3-030-13453-2_7.

Kuhn, Max and Kjell Johnson (2013). *Applied Predictive Modeling*. New York, NY: Springer New York. ISBN: 978-1-4614-6848-6 978-1-4614-6849-3. DOI: 10.1007/978-1-4614-6849-3.

Kumari, Rashmi (2021). "Adaptness Assessment of Pedestrian Street during Crises like Covid-19". In: *Journal of Urban and Environmental Engineering*. DOI: 10.4090/juee.2021.v15n1.050057.

Kyurkchiev, Nikolay and Svetoslav Markov (2015). *Sigmoid Functions: Some Approximation and Modelling Aspects Some Moduli in Programming Environment MATHEMATICA*. 1. Aufl. Saarbrücken: LAP LAMBERT Academic Publishing. ISBN: 978-3-659-76045-7.

Lam, Siu Kwan, Antoine Pitrou, and Stanley Seibert (2015). "Numba: A Llvm-Based Python Jit Compiler". In: *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, pp. 1–6.

Laube, Patrick, Mark De Berg, and Marc Van Kreveld (2008). "Spatial Support and Spatial Confidence for Spatial Association Rules". In: *Headway in Spatial Data Handling*. Ed. by Anne Ruas and Christopher Gold. Red. by William Cartwright et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 575–593. ISBN: 978-3-540-68565-4 978-3-540-68566-1. DOI: 10.1007/978-3-540-68566-1_33.

Lazar, Seth (Apr. 2025). "Governing the Algorithmic City". In: *Philosophy & Public Affairs* 53.2, pp. 102–168. ISSN: 0048-3915, 1088-4963. DOI: 10.1111/papa.12279.

Leden, Lars et al. (Jan. 2014). "A Sustainable City Environment through Child Safety and Mobility—a Challenge Based on ITS?" In: *Accident Analysis and Prevention* 62, pp. 406–414. ISSN: 00014575. DOI: 10.1016/j.aap.2013.06.013.

Lemaître, Guillaume, Fernando Nogueira, and Christos K. Aridas (2017). "Imbalanced-Learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning". In: *Journal of Machine Learning Research* 18.17, pp. 1–5.

Leur, Paul De and Tarek Sayed (Jan. 2002). "Development of a Road Safety Risk Index". In: *Transportation Research Record: Journal of the Transportation Research Board* 1784.1, pp. 33–42. ISSN: 0361-1981, 2169-4052. DOI: 10.3141/1784-05.

Lin, Yida et al. (Dec. 4, 2024). "Deep Learning-Based Depth Map Generation and YOLO-integrated Distance Estimation for Radiata Pine Branch Detection Using Drone Stereo Vision". In: *2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. 2024 39th International Conference on Image and Vision Computing New Zealand (IVCNZ). Christchurch, New Zealand: IEEE, pp. 1–6. ISBN: 979-8-3315-1877-6. DOI: 10.1109/IVCNZ64857.2024.10794464.

Lizárraga, Carmen et al. (Feb. 7, 2022). "Do University Students' Security Perceptions Influence Their Walking Preferences and Their Walking Activity? A Case Study of

Granada (Spain)". In: *Sustainability* 14.3, p. 1880. ISSN: 2071-1050. DOI: 10.3390/su14031880.

Lundberg, Scott M and Su-In Lee (2017). "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc.

Lundberg, Scott M. et al. (2019). *Explainable AI for Trees: From Local Explanations to Global Understanding*. Version 1. DOI: 10.48550/ARXIV.1905.04610. URL: https://arxiv.org/abs/1905.04610 (visited on 10/15/2025). Pre-published.

Mapillary (2025). *Mapillary: Collaborative Street-Level Imagery Platform*.

McDonald, Noreen et al. (2014). "Costs of School Transportation: Quantifying the Fiscal Impacts of Encouraging Walking and Bicycling for School Travel". In: *Transportation*. DOI: 10.1007/s11116-014-9569-7.

Men, Jinkun et al. (May 2022). "A Pareto-Based Multi-Objective Network Design Approach for Mitigating the Risk of Hazardous Materials Transportation". In: *Process Safety and Environmental Protection* 161, pp. 860–875. ISSN: 09575820. DOI: 10.1016/j.psep.2022.03.048.

Mesfin, Tarekegn Reta and Tolossa Jote Denbi (2022). "Assessment of Pedestrian Infrastructures of Road Transport: A Case Study of Jimma Town". In: *Journal of Sustainable Development of Transport and Logistics*. DOI: 10.14254/jsdtl.2022.7-2.3.

Milam, Adam J. et al. (2013). "Risk for Exposure to Alcohol, Tobacco, and Other Drugs on the Route to and from School: The Role of Alcohol Outlets". In: *Prevention Science*. DOI: 10.1007/s11121-012-0350-x.

Moraes, André Magalhães et al. (Mar. 17, 2025). "Effectiveness of YOLO Architectures in Tree Detection: Impact of Hyperparameter Tuning and SGD, Adam, and AdamW Optimizers". In: *Standards* 5.1, p. 9. ISSN: 2305-6703. DOI: 10.3390/standards5010009.

Morrongiello, Barbara A and Jennifer Lasenby-Lessard (Feb. 2007). "Psychological Determinants of Risk Taking by Children: An Integrative Model and Implications for Interventions: Figure 1". In: *Injury Prevention* 13.1, pp. 20–25. ISSN: 1353-8047, 1475-5785. DOI: 10.1136/ip.2005.011296.

Naik, Nikhil et al. (July 18, 2017). "Computer Vision Uncovers Predictors of Physical Urban Change". In: *Proceedings of the National Academy of Sciences* 114.29, pp. 7571–7576. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1619003114.

Neuhold, Gerhard et al. (Oct. 2017). *The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes*. Venice: IEEE. DOI: 10.1109/ICCV.2017.534.

Ning, Huan et al. (2021). "Sidewalk Extraction Using Aerial and Street View Images". In: *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/2399808321995817.

Ning, Huan et al. (July 3, 2022). "Exploring the Vertical Dimension of Street View Image Based on Deep Learning: A Case Study on Lowest Floor Elevation Estimation". In:

International Journal of Geographical Information Science 36.7, pp. 1317–1342. ISSN: 1365-8816, 1362-3087. DOI: 10.1080/13658816.2021.1981334.

Nisa, Ume (Aug. 1, 2024). "Image Augmentation Approaches for Small and Tiny Object Detection in Aerial Images: A Review". In: *Multimedia Tools and Applications* 84.19, pp. 21521–21568. ISSN: 1573-7721. DOI: 10.1007/s11042-024-19768-7.

Okuta, Ryosuke et al. (2017). "CuPy: A NumPy-compatible Library for NVIDIA GPU Calculations". In: *Proceedings of Workshop on Machine Learning Systems (Learningsys) in the Thirty-first Annual Conference on Neural Information Processing Systems (NIPS)*.

Oluyomi, Abiodun et al. (2014). "Parental Safety Concerns and Active School Commute: Correlates across Multiple Domains in the Home-to-School Journey". In: *International Journal of Behavioral Nutrition and Physical Activity*. DOI: 10.1186/1479-5868-11-32.

OpenStreetMap contributors (2025). *Open Street Map*. OpenStreetMap Foundation.

Osuret, Jimmy et al. (2022). "Road Safety Stakeholders' Perspectives of Risk Factors, Opportunities and Barriers for Child Pedestrians in Uganda: A Qualitative Study". In: DOI: 10.21203/rs.3.rs-2354183/v1.

Panter, Jenna et al. (2010). "Neighborhood, Route, and School Environments and Children's Active Commuting". In: *American Journal of Preventive Medicine*. DOI: 10.1016/j.amepre.2009.10.040.

Pascal Rengli and Fussverkehr Schweiz (Oct. 1, 2025). *Feedback on Automated School Route Safety Framework*. E-mail.

Paszke, Adam et al. (2017). "Automatic Differentiation in PyTorch". In.

Pedregosa, F. et al. (2011). "Scikit-Learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Peralta, Miguel et al. (2020). "Active Commuting to School and Physical Activity Levels among 11 to 16 Year-Old Adolescents from 63 Low- and Middle-Income Countries". In: *International Journal of Environmental Research and Public Health*. DOI: 10.3390/ijerph17041276.

Presence Switzerland (Apr. 24, 2024). *Föderalismus*. Föderalismus. URL: https://www.aboutswitzerland.eda.admin.ch/de/foederalismus (visited on 10/05/2025).

QGIS Development Team (2025). *QGIS Geographic Information System*. manual. QGIS Association.

Quattrone, Giovanni, Licia Capra, and Pasquale De Meo (Feb. 28, 2015). "There's No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-Sourcing Datasets". In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. CSCW '15: Computer Supported Cooperative Work and Social Computing. Vancouver BC Canada: ACM, pp. 1021–1032. ISBN: 978-1-4503-2922-4. DOI: 10.1145/2675133.2675235.

Rahman, Mohammad Lutfur et al. (2020). "A Conceptual Framework for Modelling Safe Walking and Cycling Routes to High Schools". In: *International Journal of Environmental Research and Public Health*. DOI: 10.3390/ijerph17093318.

Ramírez, Sebastián (2018). *FastAPI*.

Redmon, Joseph et al. (June 2016). "You Only Look Once: Unified, Real-Time Object Detection". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, pp. 779–788. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.91.

Rothman, Linda et al. (2013). "Walking and Child Pedestrian Injury: A Systematic Review of Built Environment Correlates of Safe Walking". In: *Injury Prevention*. DOI: 10.1136/injuryprev-2012-040701.

Rothman, Linda et al. (Sept. 2015). "Associations between Parents Perception of Traffic Danger, the Built Environment and Walking to School". In: *Journal of Transport & Health* 2.3, pp. 327–335. ISSN: 22141405. DOI: 10.1016/j.jth.2015.05.004.

Rothman, Linda et al. (2019). "Spatial Distribution of Roadway Environment Features Related to Child Pedestrian Safety by Census Tract Income in Toronto, Canada". In: *Injury Prevention*. DOI: 10.1136/injuryprev-2018-043125.

Saeedan, Majid and Ahmed Eldawy (Nov. 2022). "Spatial Parquet: A Column File Format for Geospatial Data Lakes". In: *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*. SIGSPATIAL '22: The 30th International Conference on Advances in Geographic Information Systems. Seattle Washington: ACM, pp. 1–4. ISBN: 978-1-4503-9529-8. DOI: 10.1145/3557915.3561038.

Safe2School (2025). *Luege / Brämse / Halte: Die Kampagne Für Sichere Schulwege*. Safe2school. URL: https://www.safe2school.ch/ (visited on 10/05/2025).

Sarmiento, O L et al. (2015). "Relationships between Active School Transport and Adiposity Indicators in School-Age Children from Low-, Middle- and High-Income Countries". In: *International Journal of Obesity Supplements*. DOI: 10.1038/ijosup.2015.27.

Savolainen, Peter T. et al. (Sept. 2011). "The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives". In: *Accident Analysis and Prevention* 43.5, pp. 1666–1676. ISSN: 00014575. DOI: 10.1016/j.aap.2011.03.025.

Schwebel, David C. and Leslie A. McClure (2014). "Children's Pedestrian Route Selection: Efficacy of a Video and Internet Training Protocol". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (PART A), pp. 171–179. ISSN: 13698478. DOI: 10.1016/j.trf.2014.07.005. PMID: 25170289.

Silva, Philippe Barbosa, Michelle Andrade, and Sara Ferreira (Dec. 2020). "Machine Learning Applied to Road Safety Modeling: A Systematic Literature Review". In: *Journal of Traffic and Transportation Engineering (english Edition)* 7.6, pp. 775–790. ISSN: 20957564. DOI: 10.1016/j.jtte.2020.07.004.

Simonov, Kirill and Contributors (2025). *PyYAML – YAML Parser and Emitter for Python*.

Son, Tim Heinrich et al. (July 2023). "Algorithmic Urban Planning for Smart and Sustainable Development: Systematic Review of the Literature". In: *Sustainable Cities and Society* 94, p. 104562. ISSN: 22106707. DOI: 10.1016/j.scs.2023.104562.

Stadt Zürich (2025). *Stadtplan Zürich*. Stadtplan. URL: https://www.maps.stadt-zuerich.ch/zueriplan3/stadtplan.aspx (visited on 09/06/2025).

Stadt Zürich, Tiefbauamt (2024). *Fuss- Und Velowegnetz Der Stadt Zürich*. Open Data Portal Stadt Zürich.

Stadtpolizei Zürich (2025). *Sichere Schulwege*. Sichere Schulwege. URL: https://www.stadt-zuerich.ch/de/mobilitaet/verkehrssicherheit/schulwege.html (visited on 09/06/2025).

Starzyńska-Grześ, Małgorzata B. et al. (Dec. 31, 2023). "Computer Vision-Based Analysis of Buildings and Built Environments: A Systematic Review of Current Approaches". In: *ACM Computing Surveys* 55 (13s), pp. 1–25. ISSN: 0360-0300, 1557-7341. DOI: 10.1145/3578552.

Statistik Stadt Zürich (Feb. 5, 2025). *Die Stadtzürcher Bevölkerung Wächst Weiter*. Die Stadtzürcher Bevölkerung Wächst Weiter. URL: https://www.stadt-zuerich.ch/artikel/de/statistik-und-daten/die-stadtzuercher-bevoelkerung-waechst-weiter.html (visited on 09/05/2025).

The joblib developers (Aug. 2025). *Joblib*. Version 1.5.2. Zenodo. DOI: 10.5281/zenodo.16964648.

The pandas development team (Feb. 2020). *Pandas-Dev/Pandas: Pandas*. Version latest. Zenodo. DOI: 10.5281/zenodo.3509134.

Thebault-Spieker, Jacob, Brent Hecht, and Loren Terveen (Jan. 7, 2018). "Geographic Biases Are 'Born, Not Made': Exploring Contributors' Spatiotemporal Behavior in OpenStreetMap". In: *Proceedings of the 2018 ACM Conference on Supporting Groupwork*. GROUP '18: 2018 ACM Conference on Supporting Groupwork. Sanibel Island Florida USA: ACM, pp. 71–82. ISBN: 978-1-4503-5562-9. DOI: 10.1145/3148330.3148350.

Thomson, James A. et al. (2005). "Influence of Virtual Reality Training on the Roadside Crossing Judgments of Child Pedestrians". In: *Journal of Experimental Psychology: Applied* 11.3, pp. 175–186. ISSN: 1939-2192, 1076-898X. DOI: 10.1037/1076-898X.11.3.175.

Tian, Liwei et al. (2026). "Detecting Blurriness in Medical Images Using OpenCV". In: *Proceedings of the 3rd International Conference on Internet of Things, Communication and Intelligent Technology*. Ed. by Jian Dong, Long Zhang, and Tongxing Zheng. Vol. 1366. Singapore: Springer Nature Singapore, pp. 295–305. ISBN: 978-981-96-2770-7 978-981-96-2771-4. DOI: 10.1007/978-981-96-2771-4_26.

Tiefbauamt Stadt Zürich (2024). *Fuss- Und Velowegnetz Der Stadt Zürich*. Open Data Portal Stadt Zürich.

Ultralytics (2024). *YOLOv11: Real-Time Object Detection and Segmentation*.

Ultralytics (2025). *YOLO Performance Metrics*. URL: https://docs.ultralytics.com/guides/yolo-performance-metrics (visited on 10/19/2025).

United Nations (2019). *World Urbanization Prospects: The 2018 Revision*. 2018 Revision. Population Studies 52. New York: United Nations. 126 pp. ISBN: 978-92-1-148319-2.

— (2025). *Goal 11 | Department of Economic and Social Affairs*. URL: https://sdgs.un.org/goals/goal11 (visited on 10/05/2025).

Van den Bossche, Joris (Nov. 2022). *GeoPandas: Easy, Fast and Scalable Geospatial Analysis in Python*. Zenodo. DOI: 10.5281/zenodo.7320003.

Vanwolleghem, Griet et al. (2016). "Which Socio-Ecological Factors Associate with a Switch to or Maintenance of Active and Passive Transport during the Transition from Primary to Secondary School?" In: *PLOS One*. DOI: 10.1371/journal.pone.0156531.

Verhoeven, Hannah et al. (2018). "Differences in Physical Environmental Characteristics between Adolescents' Actual and Shortest Cycling Routes: A Study Using a Google Street View-Based Audit". In: *International Journal of Health Geographics*. DOI: 10.1186/s12942-018-0136-x.

Vink, Ritchie et al. (2025). *Pola-Rs/Polars: Python Polars 1.33.1*. Version py-1.33.1. Zenodo. DOI: 10.5281/zenodo.17084198.

Virtanen, Pauli et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Völkel, Thorsten and Gerhard Weber (Oct. 13, 2008). "RouteCheckr: Personalized Multicriteria Routing for Mobility Impaired Pedestrians". In: *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*. ASSETS08: The10th International ACM SIGACCESS Conference on Computers and Accessibility. Halifax Nova Scotia Canada: ACM, pp. 185–192. ISBN: 978-1-59593-976-0. DOI: 10.1145/1414471.1414506.

Wälty, Sibylle (Jan. 2, 2021). "Greater Zurich Does Not Use Land Parsimoniously: Despite the Spatial Planning Act, Which Has Been in Force since 1980". In: *Journal of Urbanism: International Research on Placemaking and Urban Sustainability* 14.1, pp. 58–74. ISSN: 1754-9175, 1754-9183. DOI: 10.1080/17549175.2020.1762707.

Wolf Thomas and Debut, Lysandre and Sanh (Oct. 2020). *Transformers: State-of-the-art Natural Language Processing*. Version v4.25.1. Zenodo. DOI: 10.5281/zenodo.7391177.

Wong, Bonny Yee-Man, Guy Faulkner, and Ron Buliung (2011). "GIS Measured Environmental Correlates of Active School Transport: A Systematic Review of 14 Studies". In: *International Journal of Behavioral Nutrition and Physical Activity* 8.1, p. 39. ISSN: 1479-5868. DOI: 10.1186/1479-5868-8-39.

Yang, Lihe et al. (2024). *Depth Anything V2*.

Yang, Zhehui et al. (Dec. 19, 2022). "Development of a Large-Scale Roadside Facility Detection Model Based on the Mapillary Dataset". In: *Sensors* 22.24, p. 9992. ISSN: 1424-8220. DOI: 10.3390/s22249992.

Zhang, Yonglin and Rencai Dong (Mar. 2018). "Impacts of Street-Visible Greenery on Housing Prices: Evidence from a Hedonic Price Model and a Massive Street View Image Dataset in Beijing". In: *ISPRS International Journal of Geo-Information* 7.3, p. 104. ISSN: 2220-9964. DOI: 10.3390/ijgi7030104.

Zhou, Ling, Ziyang Gong, and Pengcheng Xiang (Apr. 22, 2024). "Distributed Computing and Inference for Big Data". In: *Annual Review of Statistics and Its Application* 11.1, pp. 533–551. ISSN: 2326-8298, 2326-831X. DOI: 10.1146/annurev-statistics-040522-021241.

Zito, Giuseppe A. et al. (2015). "Street Crossing Behavior in Younger and Older Pedestrians: An Eye- and Head-Tracking Study". In: *BMC Geriatrics*. DOI: 10.1186/s12877-015-0175-0.

Zürich, Stadt (2025). *Stadtplan*.

Appendix

Complete Per-Class Results (Street-Level Model)

This appendix provides the full per-class detection results for the Street-Level Model. For each class, the table lists precision, recall, and AP_{50} for both the bounding box and segmentation branches. These detailed metrics complement the summary of safety-relevant features presented in Section 5.4.

Table 1: Full per-class results for the Street-Level Model (Bounding Box and Segmentation).

Class	Bounding Box			Segmentation		
	Precision	Recall	AP_{50}	Precision	Recall	F1
animal–bird	0.55	0.22	0.24	0.56	0.22	0.32
animal–ground-animal	1.00	0.00	0.00	1.00	0.00	0.00
construction–barrier–ambiguous	0.00	0.00	0.00	0.00	0.00	0.00
construction–barrier–concrete-block	0.62	0.42	0.50	0.42	0.26	0.32
construction–barrier–curb	0.48	0.36	0.37	0.34	0.24	0.28
construction–barrier–fence	0.47	0.39	0.39	0.45	0.34	0.39
construction–barrier–guard-rail	0.48	0.48	0.48	0.43	0.42	0.43
construction–barrier–road-median	0.68	0.71	0.69	0.44	0.43	0.43
construction–barrier–road-side	0.32	0.20	0.21	0.22	0.13	0.17
construction–barrier–separator	0.15	0.04	0.13	0.00	0.00	0.00
construction–barrier–temporary	0.70	0.24	0.32	0.73	0.24	0.36
construction–barrier–wall	0.33	0.25	0.24	0.28	0.19	0.23
construction–flat–bike-lane	0.27	0.29	0.14	0.28	0.29	0.28
construction–flat–crosswalk-plain	0.43	0.33	0.35	0.40	0.26	0.31
construction–flat–curb-cut	0.44	0.10	0.11	0.35	0.06	0.11
construction–flat–driveway	0.50	0.07	0.08	0.60	0.05	0.09
construction–flat–parking	0.42	0.34	0.29	0.19	0.14	0.16
construction–flat–parking-aisle	1.00	0.00	0.00	1.00	0.00	0.00
construction–flat–pedestrian-area	0.74	0.36	0.49	0.26	0.12	0.17
construction–flat–rail-track	0.35	0.62	0.56	0.29	0.50	0.36
construction–flat–road	0.70	0.79	0.85	0.60	0.66	0.63
construction–flat–road-shoulder	0.50	0.59	0.59	0.27	0.30	0.28
construction–flat–service-lane	0.46	0.43	0.47	0.44	0.38	0.41
construction–flat–sidewalk	0.51	0.49	0.48	0.32	0.29	0.30
construction–flat–traffic-island	0.42	0.32	0.31	0.27	0.18	0.22
construction–structure–bridge	0.59	0.47	0.51	0.56	0.40	0.47

Table 1: (continued)

Class	Bounding Box			Segmentation		
	Precision	Recall	AP_{50}	Precision	Recall	F1
construction–structure–building	0.56	0.62	0.61	0.51	0.53	0.52
construction–structure–garage	1.00	0.00	0.00	1.00	0.00	0.00
construction–structure–tunnel	0.43	0.33	0.36	0.44	0.33	0.38
human–person–individual	0.51	0.38	0.39	0.41	0.29	0.34
human–person–person-group	0.14	0.01	0.04	0.04	0.00	0.00
human–rider–bicyclist	0.54	0.59	0.64	0.33	0.35	0.34
human–rider–motorcyclist	0.60	0.43	0.46	0.15	0.09	0.11
human–rider–other-rider	1.00	0.00	0.00	1.00	0.00	0.00
marking–continuous–dashed	0.55	0.45	0.48	0.39	0.29	0.33
marking–continuous–solid	0.59	0.49	0.54	0.41	0.32	0.36
marking–continuous–zigzag	1.00	0.00	0.00	1.00	0.00	0.00
marking–discrete–ambiguous	1.00	0.00	0.00	1.00	0.00	0.00
marking–discrete–arrow–left	0.34	0.14	0.26	0.37	0.14	0.21
marking–discrete–arrow–other	1.00	0.00	0.00	1.00	0.00	0.00
marking–discrete–arrow–right	0.23	0.13	0.16	0.13	0.07	0.09
marking–discrete–arrow–split	0.00	0.00	0.00	0.00	0.00	0.00
marking–discrete–arrow–straight	0.31	0.25	0.31	0.32	0.25	0.28
marking–discrete–crosswalk-zebra	0.38	0.29	0.31	0.34	0.23	0.27
marking–discrete–give-way-row	1.00	0.00	0.09	1.00	0.00	0.00
marking–discrete–hatched–chevron	0.40	0.25	0.18	0.43	0.25	0.31
marking–discrete–hatched–diagonal	0.43	0.33	0.25	0.20	0.13	0.16
marking–discrete–other-marking	0.39	0.15	0.20	0.31	0.09	0.14
marking–discrete–stop-line	0.22	0.06	0.09	0.30	0.08	0.13
marking–discrete–symbol–bicycle	0.50	0.22	0.30	0.52	0.22	0.31
marking–discrete–symbol–other	1.00	0.00	0.00	1.00	0.00	0.00
marking–discrete–text	0.37	0.28	0.33	0.41	0.28	0.33
marking-only–discrete–crosswalk-zebra	1.00	0.00	0.28	1.00	0.00	0.00
marking-only–discrete–other-marking	1.00	0.00	0.00	1.00	0.00	0.00
nature–mountain	0.38	0.35	0.30	0.20	0.15	0.17
nature–sand	1.00	0.00	0.50	1.00	0.00	0.00
nature–sky	0.82	0.99	0.98	0.79	0.92	0.85
nature–snow	0.50	0.26	0.35	0.29	0.13	0.18
nature–terrain	0.42	0.31	0.31	0.37	0.25	0.30

Table 1: (continued)

Class	Bounding Box			Segmentation		
	Precision	Recall	AP_{50}	Precision	Recall	F1
nature–vegetation	0.54	0.58	0.55	0.51	0.53	0.52
nature–water	0.31	0.27	0.33	0.21	0.17	0.19
object–banner	0.55	0.36	0.43	0.55	0.35	0.43
object–bench	0.32	0.19	0.17	0.18	0.10	0.13
object–bike-rack	0.00	0.00	0.00	0.00	0.00	0.00
object–catch-basin	0.53	0.23	0.25	0.45	0.18	0.25
object–cctv-camera	0.67	0.22	0.28	0.50	0.16	0.24
object–fire-hydrant	0.71	0.33	0.40	0.67	0.27	0.38
object–junction-box	0.38	0.26	0.21	0.43	0.26	0.33
object–mailbox	1.00	0.00	0.09	1.00	0.00	0.00
object–manhole	0.51	0.41	0.43	0.54	0.41	0.47
object–parking-meter	1.00	0.00	0.00	1.00	0.00	0.00
object–phone-booth	0.00	0.00	0.00	0.00	0.00	0.00
object–sign–advertisement	0.41	0.41	0.39	0.40	0.38	0.39
object–sign–ambiguous	1.00	0.00	0.05	1.00	0.00	0.00
object–sign–back	0.00	0.00	0.05	0.00	0.00	0.00
object–sign–information	0.25	0.14	0.16	0.25	0.11	0.15
object–sign–other	0.28	0.11	0.13	0.30	0.11	0.16
object–sign–store	0.43	0.40	0.38	0.43	0.38	0.40
object–street-light	0.75	0.35	0.45	0.69	0.31	0.43
object–support–pole	0.51	0.32	0.35	0.37	0.22	0.28
object–support–pole-group	0.57	0.03	0.10	0.00	0.00	0.00
object–support–traffic-sign-frame	0.61	0.64	0.66	0.39	0.38	0.39
object–support–utility-pole	0.46	0.51	0.49	0.42	0.43	0.43
object–traffic-cone	0.69	0.51	0.60	0.70	0.51	0.59
object–traffic-light–cyclists	0.00	0.00	0.00	0.00	0.00	0.00
object–traffic-light–general-horizontal	0.74	0.52	0.59	0.74	0.51	0.60
object–traffic-light–general-single	0.44	0.20	0.12	0.46	0.20	0.28
object–traffic-light–general-upright	0.66	0.64	0.67	0.63	0.61	0.62
object–traffic-light–other	0.00	0.00	0.00	0.00	0.00	0.00
object–traffic-light–pedestrians	0.49	0.48	0.47	0.47	0.43	0.45
object–traffic-sign–ambiguous	0.55	0.07	0.17	0.59	0.07	0.12
object–traffic-sign–back	0.52	0.34	0.38	0.48	0.29	0.36

Table 1: (continued)

Class	Bounding Box			Segmentation		
	Precision	Recall	AP_{50}	Precision	Recall	F1
object–traffic-sign–direction-back	0.50	0.42	0.41	0.47	0.38	0.42
object–traffic-sign–direction-front	0.45	0.50	0.50	0.45	0.48	0.47
object–traffic-sign–front	0.62	0.46	0.51	0.58	0.41	0.48
object–traffic-sign–information-parking	0.33	0.27	0.20	0.31	0.23	0.26
object–traffic-sign–temporary-back	1.00	0.00	0.00	1.00	0.00	0.00
object–traffic-sign–temporary-front	0.66	0.29	0.27	0.63	0.25	0.36
object–trash-can	0.34	0.38	0.40	0.34	0.36	0.35
object–vehicle–bicycle	0.45	0.46	0.43	0.39	0.35	0.37
object–vehicle–boat	1.00	0.00	0.00	1.00	0.00	0.00
object–vehicle–bus	0.37	0.65	0.60	0.40	0.65	0.50
object–vehicle–car	0.62	0.74	0.74	0.56	0.64	0.60
object–vehicle–motorcycle	0.46	0.44	0.45	0.45	0.42	0.43
object–vehicle–on-rails	0.61	0.50	0.58	0.63	0.50	0.56
object–vehicle–other-vehicle	0.25	0.10	0.27	0.20	0.08	0.11
object–vehicle–truck	0.47	0.57	0.60	0.47	0.56	0.51
object–vehicle–vehicle-group	0.35	0.16	0.21	0.27	0.10	0.14
object–vehicle–wheeled-slow	1.00	0.00	0.00	1.00	0.00	0.00
object–water-valve	0.00	0.00	0.02	0.00	0.00	0.00
void–car-mount	0.45	0.75	0.78	0.45	0.75	0.56
void–dynamic	0.40	0.17	0.16	0.44	0.18	0.26
void–ego-vehicle	0.63	1.00	0.99	0.56	0.88	0.68
void–ground	0.23	0.08	0.08	0.25	0.07	0.11
void–static	0.42	0.16	0.18	0.41	0.14	0.21

SHAP Values

This section provides the complete list of feature importance values derived from the SHAP analysis of the machine learning classifier. It complements the summary presented in the main results section and includes all model input features ranked by their mean absolute SHAP value.

Table 2: Complete SHAP feature importance values for all features.

feature	mean_abs_shap_value
rast_dashed_road_marking	0.022963
tram	0.015608
rast_traffic_lights	0.014799
crossing_with_pedestrian	0.013852
pedestrian_crossing	0.012285
rast_tram_rail	0.011773
rast_human_person	0.010831
pedestrian_crossing_without_island	0.009636
rast_tram_track	0.008818
crossing_with_tram	0.008469
footway_crossing	0.008159
rast_crosswalk	0.008073
length	0.007135
rast_traffic_island	0.006768
rast_other_traffic_sign	0.006568
rast_other_road_construction	0.006046
rast_other_object	0.005791
rast_traffic_signs	0.005611
rast_other_road_marking	0.005394
length_m	0.005277
total_length	0.005174
rast_stop_line	0.005024
car	0.004969
rast_nature	0.004731
rast_road_surfaces	0.004356
rast_motorized_vehicles_on_road	0.004233
bicycle_path	0.004096
num___crossing_with_tram_and_pedestrian	0.003950
crossing	0.003889

Continued on next page

Table 2: Complete SHAP feature importance values for all features.

feature	mean_abs_shap_value
rast_associated_with_bicycle	0.003623
num___fussgaengerstreifen_mit_insel	0.003377
arrow_marking	0.003179
num___rast_construction_parking	0.003040
num___rast_street_light	0.002888
num___crossing_without_pedestrian	0.002716
num___bushaltestelle	0.002715
cat___footway_sidewalk	0.002695
num___crossing_without_marking	0.002569
num___rast_arrow	0.002102
num___has_island_and_marked	0.002041
num___deg_v	0.001985
num___vortritt	0.001661
num___30er_zone_markierung	0.001620
cat___footway_traffic_island	0.001477
cat___footway_None	0.001211
num___deg_u	0.001051
num___rast_bridge	0.001031
num___rast_pedestrian_area	0.000896
num___30zone	0.000679
num___crossing_with_tram_has_island_and_...	0.000597
cat___highway_None	0.000573
cat___highway_footway	0.000416
num___stopp_markierung	0.000339
num___rast_tunnel	0.000321
num___crossing_with_tram_without_marking	0.000309
num___crossing_with_tram_without_pedestrian	0.000307
num___velo_markierung	0.000200
cat___highway_living_street	0.000154
cat___highway_path	0.000031
num___is_leaf	0.000005
cat___footway_link	0.000000
cat___busway_None	0.000000

Continued on next page

Table 2: Complete SHAP feature importance values for all features.

feature	mean_abs_shap_value
num___zug	0.000000
num___schule	0.000000
cat___highway_steps	0.000000
cat___highway_pedestrian	0.000000
cat___overtaking_None	0.000000
cat___motorroad_None	0.000000
cat___turn_None	0.000000
num___rast_cctv_camera	0.000000

Quarter-level Safety Score distribution

Table 3: Top 5 and bottom 5 quarters ranked by mean safety score, for ML and rule-based methods.

Quarter	Min	Max	Mean	Median	Std.Dev
Affoltern	10.8	100.0	92.9	97.4	10.8
Leimbach	4.0	100.0	92.3	97.1	10.6
Höngg	4.6	100.0	91.9	97.3	10.8
Witikon	9.0	100.0	90.5	96.4	12.3
Seebach	5.7	100.0	89.9	96.3	12.7
City	6.5	100.0	68.9	71.3	21.0
Hochschulen	5.4	100.0	73.8	77.6	19.8
Rathaus	5.1	100.0	75.1	79.5	18.9
Escher Wyss	7.3	100.0	76.3	80.5	18.1
Hard	7.6	100.0	76.8	80.8	18.3

(a) ML method

Quarter	Min	Max	Mean	Median	Std.Dev
Affoltern	0.0	100.0	91.8	95.3	13.8
Höngg	0.0	100.0	91.4	95.2	13.5
Leimbach	0.0	100.0	90.7	94.8	13.9
Witikon	0.0	100.0	89.6	94.0	14.4
Seebach	0.0	100.0	89.1	93.8	14.6
City	0.0	88.2	69.2	74.0	20.8
Hochschulen	0.0	89.3	72.4	77.0	19.9
Rathaus	0.0	90.1	74.5	79.1	19.3
Escher Wyss	0.0	91.0	75.2	79.7	18.9
Hard	0.0	89.9	75.8	80.0	18.7

(b) Rule-based method