Universität Zürich[UZH]

Geographisches Institut

Detecting Pedestrian Accessibility Features Using Deep Learning: A Case Study in the City of Zurich with Street View Imagery

GEO 511 Master's Thesis

Author: Leqi Wang, 23-744-840

Supervised by: Dr. Hoda Allahbakhshi

Faculty representative: Prof. Dr. Ross Purves

31.03.2026

Universität
Zürich[UZH]

Master's Thesis
presented to the Faculty of Mathematics and Natural Sciences
of the University of Zurich
for the degree of
Master of Science

Detecting Pedestrian Accessibility Features Using Deep Learning: A Case Study in the City of Zurich with Street View Imagery

Author: LEQI WANG
Matriculation Nr.: 23-744-840

Supervisor: Dr. Hoda Allahbakhshi
Faculty Member: Prof. Dr. Ross Purves

GEO 511 Master's Thesis
Department of Geography

Submission Date: 31.03.2026

Abstract

Rapid urbanization has made the creation of inclusive and equitable cities increasingly important, particularly for citizens with disabilities. Despite advances in urban planning, mobility-impaired pedestrians continue to face significant challenges in daily travel, including inadequate accessibility features, obstacles on pedestrian pathways, and a lack of reliable accessibility data to support accessible navigation. Previous studies have addressed data gaps through crowdsourcing, participatory surveys, and computer vision approaches, yet automated and scalable assessment of crosswalk-specific accessibility remains limited. This thesis develops an end-to-end automated pipeline for crosswalk accessibility assessment in the City of Zurich. A You Only Look Once version 11 Small (YOLO11s) object detection model is trained to detect four crosswalk-related accessibility features from street view imagery, including curb ramps, marked crossings, tram tracks, and traffic lights. The model achieves an overall mAP50 of 0.896. Marked crossings are detected with the highest accuracy at a recall of 0.983, while curb ramps present the greatest challenge with a recall of 0.769. A Crosswalk Severity Index (CSI) is subsequently developed by integrating the detected features with crosswalk length estimates derived from monocular depth estimation. The CSI is compared against the current crosswalk severity rating. The full pipeline is applied to an independent set of tram-mounted street view images to analyze the model transferability. The results demonstrate that the pipeline produces consistent and reliable feature detection and severity assessments under similar urban contexts. While the proposed approach offers a scalable, objective, and cost-effective alternative to manual surveying methods, it retains inherent uncertainties. Rather than replacing participatory approaches, cross-referencing automated detection results with human annotation could generate more reliable, accurate, and up-to-date accessibility datasets for evidence-based urban planning.

Acknowledgment

I would like to express my sincere gratitude to Dr. Hoda Allahbakhsi, the supervisor of this thesis, for her strong support and guidance. Her working style and approach to scientific research have greatly shaped my academic characteristics. Her encouragement and constructive suggestions have been valuable in guiding me towards further academic research. I wish her all the best in both her research and personal life.

I would like to thank Prof. Dr. Ross Purves, faculty member of this thesis, for his support and valuable feedback throughout the thesis. I specifically thank Ross for his patience in helping me define the research topic, and for teaching me the routines and standards of scientific research.

This thesis is embedded within the Zurich Urban Reachability & Accessibility Enhancement Through Digital Technology (ZuReach) project, funded by the Digitalization Initiative of the Zurich Higher Education Institutions (DIZH). The project builds upon and continues the city-funded Zurich Accessible City (ZuriACT) project; both projects were awarded to Dr. Hoda Allahbakhshi. The ZuriACT project provided valuable and up-to-date sidewalk spatial accessibility datasets.

I would like to express my gratitude to Alexandra-Ioana Georgescu for her cooperation and for providing preprocessed ZuriACT spatial data, as well as mobility-impaired persons' perception data on sidewalk elements, which were essential for this work.

I would like to thank iNovitas AG and its staff for their support. They provided important data sources, including access to the Infra3D platform, raw street view images, and point cloud data. Their cooperation and responsiveness were greatly appreciated.

I sincerely appreciate the support of my friends in Zurich and Munich, both mentally and personally. It has been a great honor to get to know each one of you, and my time in Zurich would not have been the same without you.

Finally, I would like to thank my parents for their unwavering support, both emotionally and financially, throughout my studies and the completion of this thesis.

Contents

List of Acronyms

AP	Average Precision
APS	Accessible Pedestrian Signals
CNN	Convolutional Neural Networks
CV	Computer Vision
GIS	Geographic Information Systems
GSV	Google Street View
IoU	Intersection over Union
LLMs	Large Language Models
LOS	Level of Service
mAP	mean Average Precision
OSM	OpenStreetMap
SDGs	Sustainable Development Goals
UN	The United Nations
UTF-8	Unicode Transformation Format (8-bit)
VR	Virtual Reality
YOLO	You Only Look Once
YOLO11s	You Only Look Once version 11 Small
YOLO11n	You Only Look Once version 11 Nano
ZuReach	Zurich Urban Reachability and Accessibility Enhancement Through Digital Technology
ZuriACT	Zurich Accessible CiTy

1. Introduction

1.1. Background and Motivation

Nowadays, the world is becoming increasingly urbanized (Hambleton, 2015), and urbanization has become the fundamental human condition (Hanson, 2004). More than 4 billion people live in urban areas, meaning over half of the world's population now resides in urban settings (Ritchie et al., 2024), underscoring the global importance of urban mobility. Spatial mobility occupies a central role in contemporary ways of life (Gallez and Motte-Baumvol, 2017), shaping how individuals access essential services, participate in social activities, and navigate the built environment on a daily basis. Under this context of rapid urbanization, ensuring that cities accommodate the full diversity of their inhabitants becomes increasingly critical. Developing inclusive cities that address these disparities aligns with Goal 11 of the 17 Sustainable Development Goals (SDGs) published by the United Nations (UN), which calls for cities and human settlements to be made inclusive, safe, resilient, and sustainable (United Nations, 2024).

Approximately 15% of the world's population live with a disability, many of whom reside in urban areas (United Nations, 2016). However, available evidence reveals a widespread lack of accessibility in built environments, from roads and housing, to public buildings and spaces (United Nations, 2016). The current built urban environment, particularly pedestrian environments, do not always respond to the needs of older people and people with disabilities (Hanson, 2004). Promoting accessible built environments, such as easy-access buildings and barrier-free sidewalks, is recognized as a key element for achieving sustainable and inclusive cities, and carries significant societal importance (Allahbakhshi, 2023). Building inclusive mobility is therefore a critical component of urban inclusiveness, where the concept of "inclusive mobility" forms part of a set of political priorities defined by several European countries to address the social dimension of transport and daily mobility policies (Gallez and Motte-Baumvol, 2017). Inclusive urban mobility and accessibility are particularly crucial for pedestrians, especially for mobility-impaired pedestrians. Mobility-impaired pedestrians' ability to navigate the urban environment is directly contingent on the quality and consistency of the pedestrian accessibility infrastructures.

Mobility impairment and mobility disability refer to a physical condition that limits a person's ability to move freely or perform activities (The Oxford Review, 2026). Mobility impairments can arise from congenital conditions, injuries, chronic illnesses, or age-related changes, and they vary widely in severity (The Oxford Review, 2026). For mobility-impaired pedestrians, such as wheelchair users and rollator users, navigating the urban environment independently presents a series of tangible physical challenges, as the current transportation systems and infrastructure have not been built to meet their requirements (Prescott, 2020). Minor infrastructural details that are unnoticed by able-bodied pedestrians can be constitute significant barriers for those relying on assistive devices. These include the lack of curb ramps at crossings, sidewalk inclinations, narrow footpaths, or recessed tram tracks (Bennett et al., 2009; Kapsalis et al., 2024). Such impediments do not merely cause inconvenience, they increase the physical difficulty of travel, introduce safety risks, and in many cases restrict independent mobility altogether (Prescott, 2020). Addressing these challenges requires a systematic approach to quantifying accessibility barriers and standardizing the evaluation of pedestrian infrastructure.

Sidewalks are important components of any transportation network and should provide inclusive walking conditions for all users (Aghaabbasi et al., 2018). At intersections, crosswalk infrastructure plays a particularly important role in pedestrian safety. Crosswalks represent the most critical and frequently encountered transition points of the pedestrian network, serving as the primary interface between mobility-impaired pedestrians and vehicular traffic. The complexity and condition of a crosswalk directly influences the safety and passability of the crossing experience for mobility-impaired users. The motivation for this thesis therefore lies in developing a data-driven understanding of crosswalk complexity and severity. The crosswalk severity is defined as quantitative measure of the difficulty and complexity experienced by mobility-impaired pedestrians when crossing a road from one side to the other. By systematically detecting and quantifying accessibility features through automated image-based sensing, this thesis aims to provide an objective and reproducible basis for assessing crosswalk accessibility. The pipeline will ultimately contribute to the creation of a safer and more inclusive pedestrian environment for mobility-impaired pedestrians in urban settings.

1.2. Research Gap

Participatory surveys and crowdsourcing-based methods are largely utilized in current studies to collect and assess pedestrian accessibility features. Both of the methods have limitations in spatial scalability, efficiency, reproducibility, and transferability (Saha et al., 2019; Panta et al., 2019). Deep learning-based computer vision methods have demonstrated promising results in automating the detection of sidewalk accessibility features from street view imagery (Weld et al., 2019). However, these approaches have remained broadly focused on general sidewalk conditions rather than the specific infrastructural complexity of crosswalks. Furthermore, the subsequent use of automated detection outputs has not been sufficiently explored in the existing literature. Specifically, more research is needed to derive a quantitative, crosswalk-specific severity indicator that explicitly reflects the accessibility needs and safety of mobility-impaired pedestrians. This gap is particularly relevant in the urban context of the City of Zurich, Switzerland. As a classic European city, the city center of Zurich is characterized by a complex pedestrian environment, including active tram corridors, high traffic volumes, and dense pedestrian activity (Zürich Tourism, 2024). This urban complexity introduces additional barriers to safe crossing for mobility-impaired pedestrians, highlighting the importance of accessible and well-equipped crosswalks. This thesis therefore aims to address these gaps by developing an automated, deep learning-driven pipeline for crosswalk accessibility feature detection and severity assessment. This thesis contributes a scalable and objective methodology for crosswalk accessibility assessment, supporting the development of safer and more inclusive urban environments for mobility-impaired pedestrians.

1.3. Research Objectives and Questions

The primary goal of this research is to develop a deep learning-based framework for detecting accessibility features and accessing severity of pedestrian crosswalks, specifically for mobility-impaired pedestrians.

The research objectives of the thesis are:

- **Automate Pedestrian Accessibility Feature Detection**: Investigate the effectiveness of deep learning and computer vision in identifying and detecting accessibility features that influence mobility of mobility-impaired pedestrians, including curb ramps, traffic lights, tram tracks, and marked crossings.

- **Develop a Crosswalk Severity Index (CSI)**: Establish a multi-criteria severity indicator that quantifies crosswalk complexity by integrating both detected accessibility features and estimated geometric measurements. The CSI is providing an objective and reproducible basis for assessing crosswalk accessibility for mobility-impaired pedestrians.

- **Evaluate Pipeline Transferability**: Apply the full pipeline of feature detection and indicator computation to an independent dataset, assessing the reproducibility and transferability of the methodology across different data sources.

The research questions of the thesis are:

- **RQ1**: How effectively can deep learning models detect pedestrian accessibility features from street view imagery?

- **RQ2**: To what extent does the proposed Crosswalk Severity Index (CSI) reflect crosswalk complexity for mobility-impaired pedestrians, and how does it compare to existing perception-based severity ratings?

- **RQ3**: How transferable is the developed pipeline to an independent street view dataset captured under different conditions?

1.4. Thesis Structure

This Master's Thesis is structured into six chapters. Following this Introduction section, Chapter 2 (Related Work) outlines the scientific context, reviews relevant studies, and highlights the research gaps. Chapter 3 (Materials and Methodology) provides the data sources and the technical framework. The framework encompasses data pre-processing, computer vision-based feature detection, crosswalk severity index computation, and the transferability analysis. Chapter 4 (Results) illustrates the outcome of the entire methodology pipeline, supported by corresponding quantitative tables and spatial visualizations. Chapter 5 (Discussion) interprets the results, compares the findings to related studies, critically examines the limitations of the study, and outlines directions for future work. The thesis conclude with Chapter 6 (Conclusion), summarizes the key findings, addresses the research questions, and draws a conclusion of the thesis.

2. Related Work

The literature on pedestrian navigation highlights a fundamental disconnect between standard urban data and the lived reality of mobility-impaired pedestrians (Völkel et al., 2008). Traditional metrics of walkability have historically focused on realistic qualities for the general population, such as route directness or connectivity. These approaches fail to account for the specific impediments encountered by individuals with visual, aural, or physical impairments (Völkel et al., 2008). These pedestrians face a range of structural difficulties, such as steep slopes, stairs, and high curbs that can render a seemingly viable path completely inaccessible to wheelchair users (Völkel et al., 2008). Beyond physical accessibility, safety and orientation also present major hurdles. For instance, visually impaired pedestrians often struggle with street crossings that lack tactile guide strips or acoustic signals, creating a high risk of accidentally entering the roadway (Li, 2021). Furthermore, the pedestrian experience is heavily influenced by social and affective factors, such as the psychological distress caused by unwanted attention or the lack of consideration from others (Sundling and Jakobsson, 2023). These challenges can significantly degrade the quality of urban mobility for mobility-impaired pedestrians.

A critical barrier to improving these experiences is the pervasive lack of high-resolution geospatial data tailored to the needs of the mobility-impaired (Völkel et al., 2008). Most current navigation systems utilize map data optimized for vehicular traffic, which lack essential micro-scale information such as pavement conditions, the exact location of lowered curbs, or the availability of accessible traffic signals (Jeong et al., 2018). Standard datasets also typically omit subjective safety ratings and temporal information, which are vital for mobility-impaired pedestrians. These pedestrians may prioritize a safer, quieter route over a shorter one or must account for temporary obstacles like construction sites or snow. Because the financial burden of collecting this granular data is often beyond the reach of public authorities, recent research suggests a shift toward collaborative, multi-modal annotation. For example, the approach by Jeong et al. (2018) enables users to share real-time observations and personal ratings, allowing for a more dynamic and personalized understanding of the pedestrian environment that incorporates both pragmatic and hedonic qualities.

2.1. Geospatial Data Acquisition of Pedestrian Accessibility Features

The significance of sidewalk infrastructure extends far beyond urban aesthetics, serving as a fundamental pillar of equitable mobility and social integration. For individuals with physical disabilities or age-related limitations, accessibility features such as curb ramps and tactile pavings are the prerequisites for independent mobility and basic social participation (Aghaabbasi et al., 2018; Asadi-Shekari et al., 2013). Capturing the presence, quality, and spatial distribution of these accessibility features is therefore essential for evidence-based urban planning and infrastructure investment. Accurately capturing the accessibility conditions of urban pedestrian environments requires reliable and spatially comprehensive data. A variety of approaches have been developed to acquire such data. These approaches include large-scale crowdsourcing platforms, participatory surveys, and automated detection methodologies. While these methodologies are widely implemented, each offers trade-offs. The following section reviews the key platform and approaches that have been employed for geospatial data acquisition in the context of inclusive urban planning. The review will highlight their respective strengths and limitations in terms of spatial coverage, data quality, and scalability.

2.1.1. Crowdsourcing

Due to the diversity of urban contexts across the world, conducting field surveys to collect pedestrian accessibility feature data require considerable economic resources and human effort (Saha et al., 2019). Web-based crowdsourcing has therefore emerged as an important and scalable alternative for collecting and validating accessibility feature data at a broad geographic scale (Estellés-Arolas and González-Ladrón-de Guevara, 2012; Panta et al., 2019).

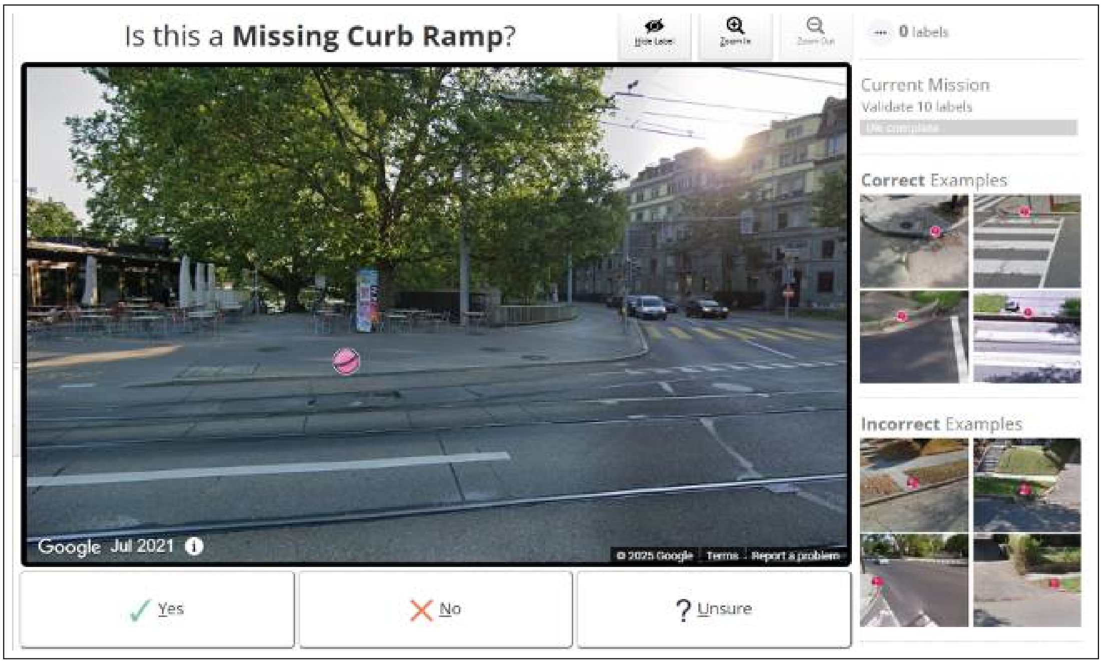
Project Sidewalk is an innovative, web-based crowdsourcing tool designed to collect sidewalk accessibility data at a massive scale by leveraging Google Street View (GSV) imagery (Saha et al., 2019). As a data acquisition platform, Project Sidewalk enables users to perform ”virtual audits” of the urban environment, allowing anyone with an internet connection to remotely label and validate accessibility features and barriers. The tool employs a mission-based interface and interactive on-boarding to guide volunteers in identifying specific infrastructure elements, such as curb ramps, missing curb ramps, surface problems, and obstacles in the pedestrian path (Saha et al., 2019). In other words, compared to a field survey, the crowdsourcing tool allows data collection without professional education (Chowdhury et al., 2016). With this crowdsourcing audit process, Project Sidewalk overcomes the high costs and logistical challenges associated with traditional field surveys, providing a cost-effective method for generating wide coverage, geo-referenced datasets of the pedestrian network.

The interface of Project Sidewalk is shown in Figure 1, where Figure 1a illustrates the labeling process and Figure 1b as the validating process. The web-based tool enables online crowd-workers to remotely label pedestrian-related accessibility problems by virtually walking through city streets in GSV (Saha et al., 2019). In an 18-month deployment study of Saha et al. (2019), 797 online users contribute 205,385 labels and audit 2,941 miles of Washington DC streets. The primary color-coded label types includes curb ramps, missing curb ramps, obstacles, surface problems, and no sidewalk (Saha et al., 2019). The users are also asked to rate the problem severity on a 5-point scale where ”5” represents an impassable barrier for someone in a manual wheelchair (Saha et al., 2019). Other than initial labeling, the platform of Project Sidewalk allows crowd-workers to validate the label (see Figure 1b). Referencing to correct and incorrect examples, the users mark ”Yes”, ”No”, or ”Unsure” for accessibility feature labels on GSV. With availability of GSV and OpenStreetMap (OSM), Project Sidewalk platform would ostensibly work globally in terms of collecting accessibility data (Saha et al., 2019).

2.1.2. Participatory Survey

In 2023, the Zurich Accessible City (ZuriACT) project was launched to map sidewalk accessibility features in District 1 of Zurich. The ZuriACT project represents a shift from large-scale online auditing to small-scale participatory and multi-modal data collection approach. Inspired and collaborated with Project Sidewalk, ZuriACT use Project Sidewalk tool for virtual audits by participants. However, the platform of Project Sidewalk has multiple limitations when applying locally, including the lack of tools in collecting measurements (Allahbakhshi, 2023). Another limitation is that Project Sidewalk has its network highly dependent on GSV and OSM, leading to lack of coverage in the entire street network of the study area of ZuriACT (Allahbakhshi, 2023; Saha et al., 2019). To accommodate these challenges, ZuriACT alternatively integrate and utilize the Infra3D web-based tool, a platform based on up-to-date and complete Street View Image (SVI) data taken by the Swiss company iNovitas (Allahbakhshi, 2023; iNovitas AG, 2023). The ZuriACT project used crowd-sourcing web-based tool to conduct participatory surveys, allowing mobility-impaired pedestrians to

(a)



(b)

Figure 1. Project Sidewalk interface: (a) the labeling interface showing accessibility feature categorization and severity rating, (b) the validation interface. Source: projectsidewalk.org

label and validate accessibility feature through the online platform. On top of Project Sidewalk, ZuriACT collect data on measurements such as sidewalk incline and width, and additional features such as pedestrian signal and occlusion (Allahbakhshi, 2023). The ZuriACT dataset includes over 9,000 data points capturing sidewalk accessibility elements such as curbs, crosswalks, surface issues, obstacles, etc. (Allahbakhshi, 2023). Building on ZuriACT, the ongoing ZuReach (Zurich Urban Reachability & Accessibility Enhancement Through Digital Technology) project extends the study area to the entire city of Zurich (Allahbakhshi, 2025). The ZuReach project aims to provide a scalable and regularly updated database containing comprehensive sidewalk accessibility information in Zurich (Department of Geography, University of Zurich, 2024).

Similarly, (Li et al., 2024) examines the potential of community-driven digital civics for sidewalk accessibility assessment through a deployment study of Project Sidewalk. The study adopts a two-way participatory approach, involving community members in the full research process, from problem identification and data collection to analysis and conclusion. This approach empowers participants to articulate local accessibility issues and advocate for infrastructure improvements in their communities. The findings demonstrate that web-based crowdsourcing tools can simultaneously serve as data collection platforms and community engagement tools. However, the quality and coverage of the data remain dependent on the level of participation, reinforcing the scalability limitations inherent to participatory approaches.

2.1.3. Deep Learning

Beyond the foundational work of Saha et al. (2019), Project Sidewalk has been deployed across multiple cities to collect thousands of accessibility labels. Studies further demonstrate the utility of Project Sidewalk as a large-scale data acquisition platform for pedestrian infrastructure mapping. Researchers have leveraged this crowdsourced data as a training and validation resource for automated detection systems. For instance, Weld et al. (2019) employed a modified residual neural network (ResNet) to automatically label sidewalk accessibility issues and validate the accuracy of community-contributed labels. With more than 300,000 image-based sidewalk accessibility labels, Weld et al. (2019) presents examination of deep learning to automatically assess sidewalks in GSV panoramas. This work demonstrates how Project Sidewalk functions not only as a repository of urban information but also as a critical training resource for computer vision models.

The growing availability of large-scale crowdsourced datasets has accelerated the adoption of deep learning methods for automated pedestrian accessibility assessment. Modern computer vision pipelines leverage Convolutional Neural Networks (CNNs) and transformer-based architectures to analyze high-resolution street-level imagery from platforms like GSV or dedicated mobile mapping systems (Spaias et al., 2025; Lawal and Agbaje, 2026). These automated approaches offer a scalable alternative to traditional field surveys, showing the possibility to generate comprehensive and continuously up-datable geo-referenced accessibility feature datasets. The combination of crowdsourcing and deep learning approach allows for a dynamic understanding of the pedestrian environment that reflects actual street conditions rather than idealized design specifications, providing a scalable solution for monitoring urban accessibility (Weld et al., 2019).

Other computer vision techniques are also utilized to detect sidewalk features in previous studies. The You Only Look Once (YOLO) framework revolutionized the field of computer vision by framing object detection as a single regression problem, where a unified CNN simultaneously predicts bounding boxes and class probabilities directly from full images in a single pass (Khanam and Hussain, 2024). The YOLO architecture is particularly suitable for identifying pedestrian accessibility features due to its proven efficacy in detecting complex objects like pedestrians in diverse, real-time environments. Research has demonstrated that specialized versions like YOLO-R can mitigate the loss of fine-grained pedestrian information by using Passthrough layers to link shallow-layer features with deep-layer information (Dow et al., 2020). This can significantly improve detection accuracy and reduce false detection rates (Dow et al., 2020). Furthermore, YOLO's high fault tolerance allows it to recognize partially obscured pedestrians and navigate challenges such as varying light conditions or complex backgrounds that often baffle traditional schemes like Haarcascade or HOG (Lan et al., 2018). By integrating YOLO with specialized pre-processing techniques, these systems can efficiently monitor urban intersections to enhance safety and provide essential geo-referenced data for individuals with mobility impairments (Lan et al., 2018).

Building on the advancements of real-time monitoring, Kandavel et al. (2025) proposes an AI-driven traffic monitoring framework to enhance urban safety and infrastructure maintenance. This study integrates YOLOv8 for high-precision object detection with Large Language Models (LLMs) (Kandavel et al., 2025). The system utilizes an anchor-free detection mechanism and decoupled heads to effectively identify overlapping objects. It achieves a 96% detection accuracy for road entities such as vehicles and pedestrians. A distinctive feature of this methodology is the incorporation of Few-Shot Learning (FSL). It allows the model to adapt to new pedestrian behaviors and unseen crossing patterns with limited labeled data, reducing false negatives by 22% compared to standard computer vision models. Furthermore, the framework leverages LLaMA 3.2B for contextual analysis, enabling the system to process detected anomalies and provide real-time maintenance recommendations with an impressively low inference latency of 75.1 ms ((Kandavel et al., 2025)).

The latest iteration, YOLOv11, introduces significant architectural enhancements designed to maximize feature extraction while maintaining computational efficiency. Key innovations include the C3k2 block, which utilizes smaller kernel sizes for faster processing. The C2PSA component integrates a spatial attention mechanism, allowing the model to focus more effectively on critical regions of an image (Khanam and Hussain, 2024). Alongside the Spatial Pyramid Pooling - Fast (SPPF) block, these advancements enable YOLOv11 to capture nuanced details across multiple scales, achieving superior mean Average Precision (mAP) and faster inference speeds compared to previous generations such as YOLOv8 and YOLOv10 (Khanam and Hussain, 2024). The progression across YOLO iterations reflects a continuous improvement in detection capability and architectural efficiency. YOLO-R's preservation of fine-grained features, YOLOv8's adaptability through Few-Shot Learning, and YOLOv11's enhanced accuracy and processing speed each contribute distinct strengths to the object detection landscape (Kandavel et al., 2025; Lan et al., 2018; Ultralytics, 2024). Together, these advancements provide a robust technological foundation for a automated deep learning pipeline, supporting the detection of crosswalk-level features and the subsequent computation of accessibility severity indicators at urban scale.

2.1.4. Crosswalk-related Accessibility Feature Detection

Crosswalk is an important accessibility feature, and can be critical for mobility-impaired pedestrians. Crosswalks represent the most critical and high-risk segments of the pedestrian network, as they facilitate the interface between foot traffic and potentially hazardous motorized vehicles Ahmed et al. (2021). Detection algorithms for these features must overcome significant environmental variability, including faded paint markings, diverse material patterns, adverse weather conditions, and frequent occlusion by parked cars or moving traffic. Recent advancements have transitioned from simple bounding-box detection to sophisticated semantic segmentation, which allows for the precise extraction of crosswalk geometry and boundaries (Soares Müller et al., 2023). This pixel-level data is vital for assessing the functional ”walkability” of a crosswalk, as it enables the automated calculation of crossing distances and the identification of pedestrian refuge islands. Furthermore, identifying the presence and condition of ancillary features such as Accessible Pedestrian Signals (APS) or tactile warning strips within these detected zones is critical for determining the safety level for visually or mobility-impaired users who rely on these cues to navigate traffic safely (Soares Müller et al., 2023).

Previous studies have utilized various approach to specifically detect crosswalks and intersections. Zhang et al. (2022) proposes the crosswalk detection network (CD-Net) based on YOLOv5 to achieve fast and accurate crosswalk detection under the vision of the vehicle-mounted camera. They implement real-time detection on Jetson nano device. This study achieves an average F1 score of 94.83% in the above complex scenarios, and an average F1 score exceeds 98% under better weather condition (Zhang et al., 2022). The study demonstrates a specific application of artificial neural network algorithm optimization methods on edge computing devices (Zhang et al., 2022). Similarly, another study by Lu et al. (2024) introduces a real-time crosswalk detector called X-CDNet based on YOLOX, expanding the model's receptive field without the addition of extra inference time. X-CDNet achieves a 93.3 AP50 and a real-time detection speed of 123 FPS in real-time crosswalk detection (Lu et al., 2024). Other than real-time detection, deep learning was also implemented in image processing to detect crosswalks and intersections. Tümen and Ergen (2020) proposes an image processing and deep learning approach for the classification of road separations, intersections, and crosswalks. The study evaluates multiple CNN architectures, including VggNet-5, LeNet, and AlexNet, alongside a newly proposed CNN model designated RoIC-CNN. As their results, RoIC-CNN achieves the highest intersection detection accuracy of 100%, outperforming VggNet-5 (93.33%), LeNet (96.67%), and AlexNet (96.67%) (Tümen and Ergen, 2020). This study demonstrates the effectiveness of the proposed architecture for road feature classification tasks.

2.1.5. Transferability Analysis

Previous literature studies and demonstrates the availability and suitability of multiple approaches in retrieving and detecting pedestrian accessibility features under urban context. As most of these approaches and methodologies are applied on certain study areas and cases, the transferability of these approaches should be further assessed. Transferability analysis evaluates the generalization of these models to ensure they remain robust when scaled to new environments. Duan et al. (2022) examines the transferability of crowdsourced data in automatically classifying sidewalk accessibility features from streetscape imagery. Two experiments are conducted: the first compares the classification performance of validated crowdsourced labels against larger but noisier aggregate datasets, while the second investigates whether labeled data collected in one city could serve as effective training data for a model deployed in another (Duan et al., 2022). The study identifies key challenges in transferring crowd-assisted accessibility assessment pipelines across diverse urban contexts. As demonstrated by Duan et al. (2022), models trained on sidewalk data in one city can potentially be transferred effectively to others, but their transferability is contingent on the degree of similarity between the source and target urban environments.

2.2. Crosswalk Accessibility Assessment

The detection and mapping of pedestrian accessibility features provides the foundational data layer upon building accessibility assessment frameworks. However, identifying the presence and location of individual features is insufficient to characterize the overall accessibility experience of a pedestrian crosswalk. A meaningful assessment requires the integration of multiple physical and infrastructural attributes into a coherent measure of crossing complexity or severity. Translating raw detection outputs into such a measure necessitates dedicated evaluation frameworks. The following section reviews existing approaches to crosswalk accessibility assessment, examining how previous studies have utilized accessibility as a quantifiable metric in crosswalk-specific severity assessment.

Ahmed et al. (2021) introduces the Pedestrian Crossing Level of Service (PCLOS) framework as a quantitative method for assessing crosswalk conditions. The framework assesses crossings across 17 indicators grouped into four categories: geometry, traffic characteristics, environment, and pedestrian behavior. Indicator weights and scores are assigned through a combination of field observations and participatory surveys. The provision of a zebra crossing is identified as the most critical indicator, while drainage near crosswalks is considered the least important. The framework produces a letter-grade service level from A to F, enabling planners to systematically identify deficiencies and prioritize infrastructure improvements (Ahmed et al., 2021). While PCLOS provides a valuable basis for multi-criteria crosswalk evaluation, it remains oriented towards general pedestrian safety rather than the specific accessibility needs of mobility-impaired users (Soares Müller et al., 2023).

The approach presented by Kadali and Vedagiri (2015) assesses the importance of land-use criteria in pedestrian Level of Service (LOS). In this framework, pedestrian perceived LOS are collected with respect to different land-use type such as shopping, residential and business areas. The study concludes that perceived safety, crossing difficulty, land-use condition, number of vehicles encountered, median width and number of lanes have significant effect on pedestrian perceived LOS at unprotected crosswalks in mixed traffic scenario. Their findings indicate that land-use context significantly alters pedestrian expectations and behaviors. For instance, pedestrians in shopping and residential areas tend to perceive lower LOS scores (averaging 4.6, or LOS E/F) as they expect higher levels of comfort compared to those in industrial or business zones (Kadali and Vedagiri, 2015). While this qualitative approach remains limitation in scalability and transferability, this study underscores that pedestrian perception is a meaningful and safety-relevant dimension of crosswalk accessibility assessment.

Crosswalk markings installed at the vehicle-pedestrian intersections are not only facilities for pedestrian, but also a function to warn drivers to watch out for pedestrians (Bian et al., 2020). Rather than assessing crosswalks from the pedestrian perspective, Bian et al. (2020) assesses crosswalk markings from the drivers' point of view, accessing the effectiveness of the newly designed crosswalk markings in China. By collecting data through driving simulator, nine indicators representing vehicle operating data, drivers' maneuvering data, and drivers' subjective evaluation are proposed. The assessment result shows that, for intersections with high or low pedestrian flow, the comprehensive effectiveness and influences on drivers' driving behaviors with presence of newly designed crosswalk markings are better than the standard markings (Bian et al., 2020). This study introduces an alternative perspective in crosswalk evaluation, highlighting that a comprehensive severity assessment framework may benefit from incorporating driver-side parameters alongside pedestrian-focused indicators.

3. Materials and Methodology

3.1. Overview

The main methodology of this thesis includes four parts: data pre-processing, computer vision, Crosswalk Severity Index calculation, and transferability analysis. A general workflow is shown in Figure 2.

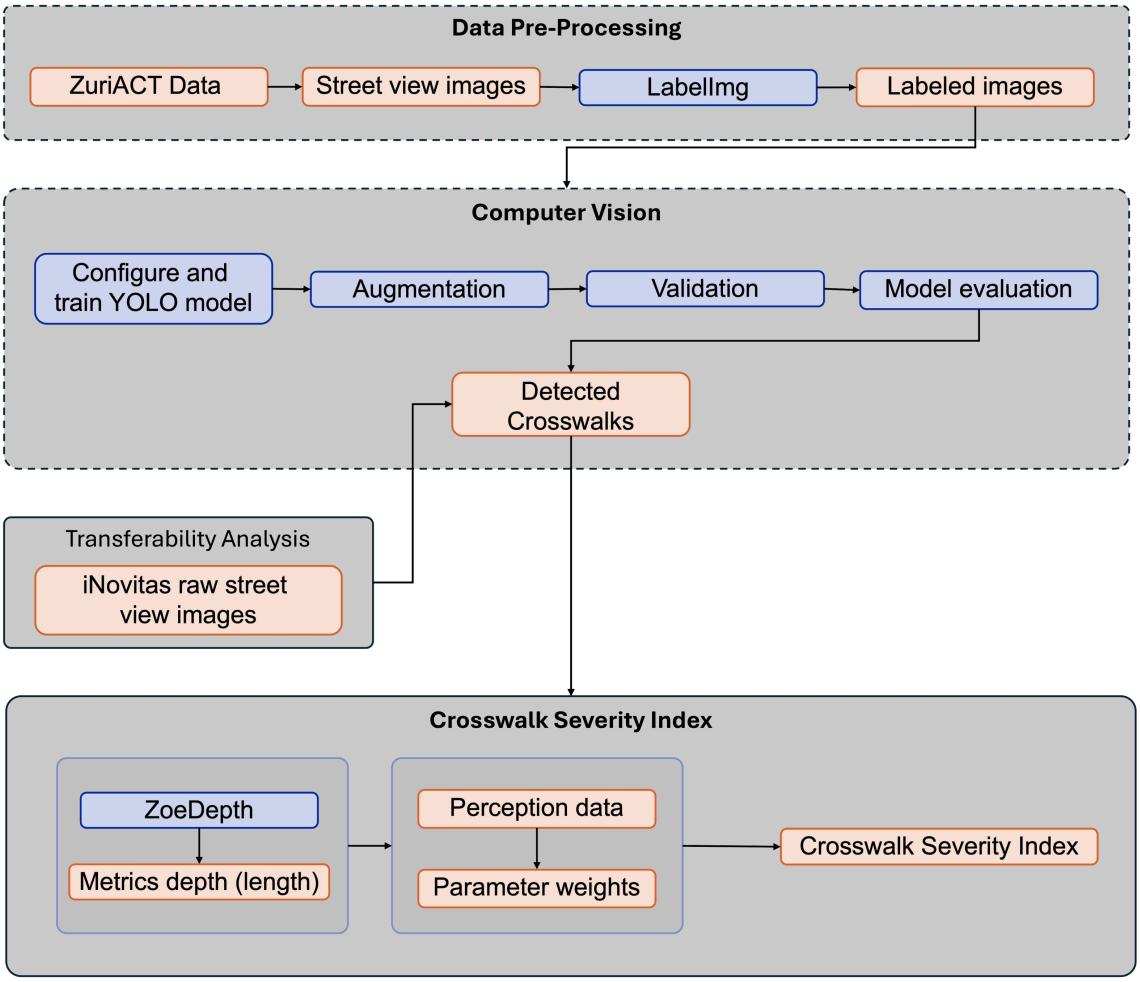


Figure 2. Workflow. Blue: tools & models. Orange: data & results.

3.2. Study Area

The main study area of the thesis is District 1 of the City of Zurich, Switzerland (Figure 3). District 1 is the most central district of Zurich, having historical Old Town on both banks of the River Limmat, it is also the commercial core of the city (Zürich Tourism, 2024). It is home to the prestigious Bahnhofstrasse, a major shopping avenue characterized by the integration of high pedestrian volumes and active tram corridors (Zürich Tourism, 2024). With this dense urban fabric, District 1 presents a complex array of street layouts and multi-modal intersections. District 1 of Zurich is a critical urban environment to access and evaluate pedestrian accessibility features, especially in the case of evaluating complex crosswalks. The geographic scope of the thesis aligns with the ZuriACT project. The transferability analysis radiates to several streets outside of District 1, including Kasernenstrasse, Stauffacherstrasse, Feldstrasse, and Hohlstrasse (see Figure 4).

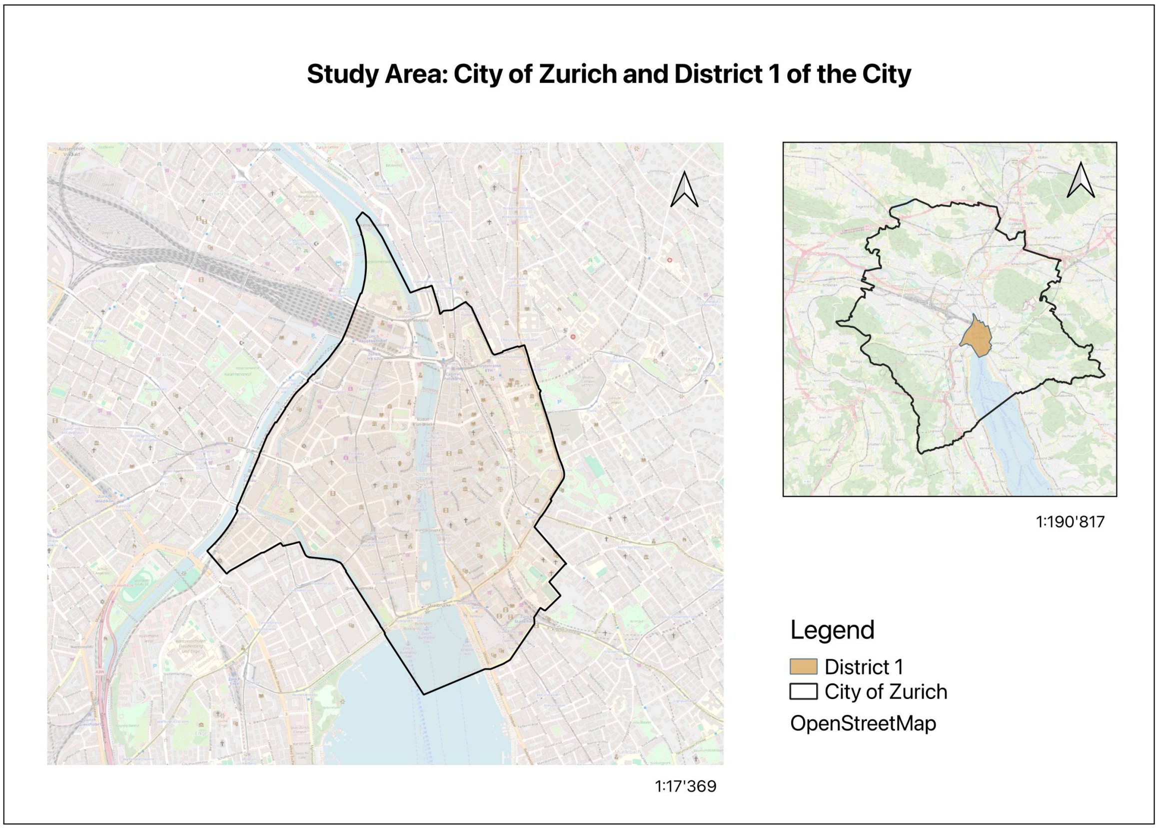


Figure 3. Map of the study area in District 1, Zurich

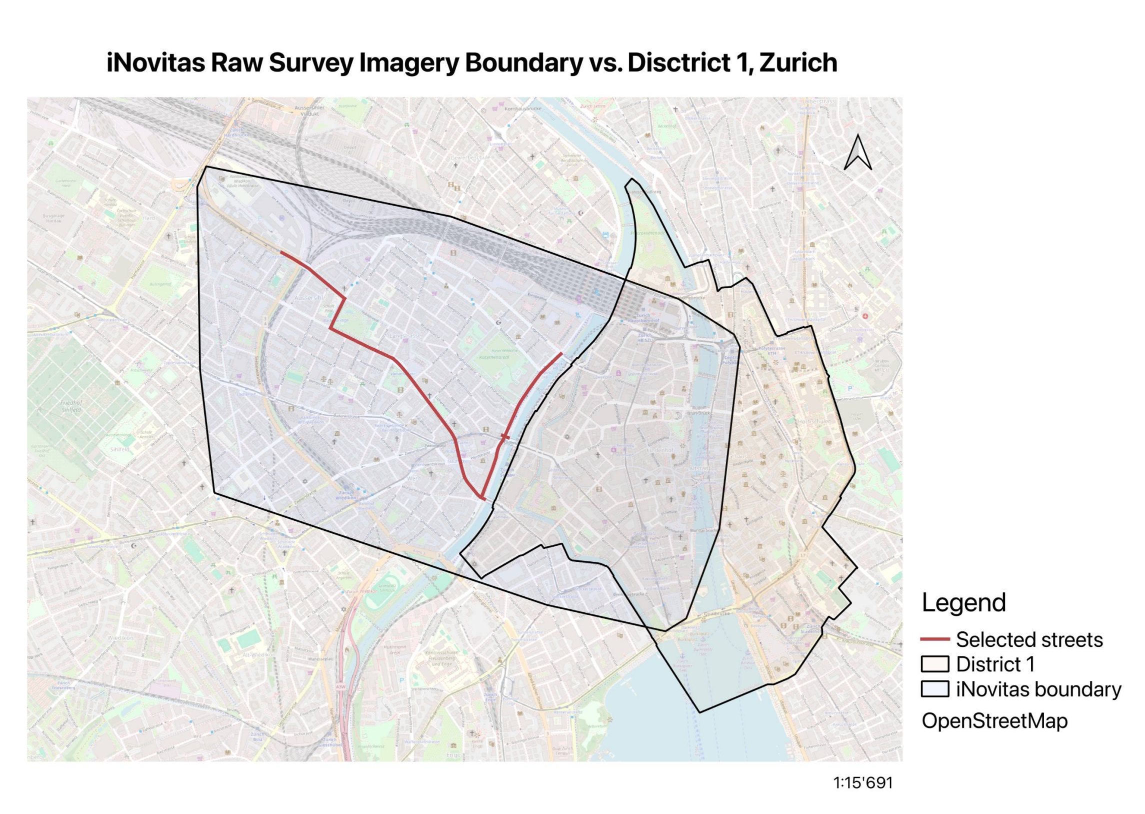


Figure 4. Geographical boundary of iNovitas raw street view image (1.4 km^2, 2024) with selected streets.

3.3. Data

The methodology of this thesis relies on the integration of two categories of data: geo-referenced point data of pedestrian accessibility features and their corresponding street view images. The data source summary is listed in Table 1 , details will be introduced in the following sections.

Table 1. Data source overview

Dataset	**Source**	**Area**
Validated crosswalk & curb ramps	ZuriACT	District 1
Validated zebra crossing	Stadt Zürich	Entire city
Street view images (screenshots)	Infra3D (iNovitas AG, 2026b)	District 1
Street view images (received)	Infra3D (iNovitas AG, 2026a)	District 1 & additional streets

3.3.1. ZuriACT

The primary source for geo-referenced pedestrian accessibility features data is the ZuriACT project. The raw dataset was published on the Open Data Zurich portal in 2024, and it provides a comprehensive inventory of sidewalk accessibility elements (Allahbakhshi, 2023; Allahbakhshi and Ardüser, 2024). While the full raw dataset contains over 9,000 data points, including surface problems, obstacles, and pedestrian signals (Allahbakhshi, 2023), this thesis specifically utilizes the point data for crosswalks and curb ramps (Figure 5). These datasets are initially collected by participants using image-based digital tools. To ensure data integrity, each feature is validated by at least two experts and subsequently clustered to produce unique, single-point representations for every crosswalk and curb ramp (Allahbakhshi et al., 2026).

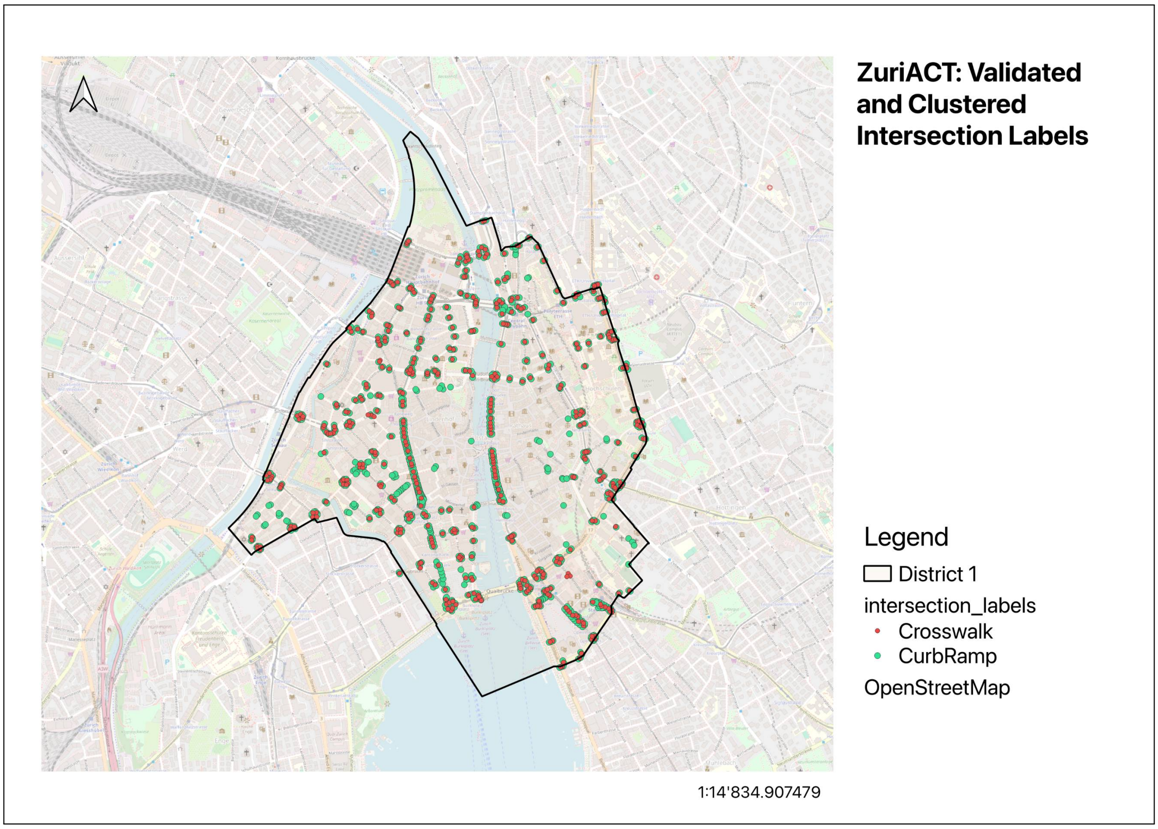


Figure 5. Validated and clustered crosswalks and curb ramps in District 1, Zurich

While the initial ZuriACT project focuses exclusively on District 1 of the City of Zurich, the follow-up project, ZuReach (Zurich Urban Reachability and Accessibility Enhancement Through Digital Technology), extending the study area to encompass the entire city (Allahbakhshi, 2025). In previous PhD and MSc thesis of Georgescu et al. (2026) and Ballinari (2025), perception surveys are conducted with 30 mobility-impaired participants to evaluate pedestrian accessibility features. Participants are asked to rate various urban elements based on their real-life experience. For barriers, the scale ranged from 1 to 5, where 1 is fully passible and 5 as not passable. For facilitators, the scale ranged from 1 to 5 where 1 is very facilitating and 5 as not facilitating. The specific elements assessed in the survey includes:

- **Barriers:**
  - High curbs
  - No zebra crosswalk
  - No accessible pedestrian signals
  - Tram tracks at crosswalk
  - Long crosswalk
  - Crosswalk surface problems
  - Fading paint at crosswalk
- **Facilitators:**
  - Lowered curbs
  - Marked zebra crosswalks
  - Presence of accessible pedestrian signals

These perception data, stored in a CSV file, provides the empirical basis for calculating the weights used in the Crosswalk Severity Index (CSI). By utilizing the mean scores from this diverse group of mobility-impaired individuals, the thesis ensures that the final automation pipeline prioritizes infrastructure improvements based on real-world user needs. These perception data participate a quantitative foundation to the thesis, anchoring the qualitative experiences of mobility-impaired individuals into measurable indices.

3.3.2. Infra3D Street View Images

High-resolution visual data is sourced from Infra3D, a specialized platform for digitized road and railway infrastructure provided by the Swiss service provider iNovitas AG (iNovitas AG, 2026b). The Infra3D database provides detailed, georeferenced 3D representations of Swiss streetscapes with centimeter-level precision, providing an exceptionally high degree of spatial accuracy for District 1 of the City of Zurich (iNovitas AG, 2026a).

Compared to conventional services such as Google Street View (GSV), Infra3D provides several distinct advantages for the cases of Swiss cities. The imagery is more up-to-date, more complete, more accurate, and more systematically organized. Other than standard road surveying in GSV, a unique feature of this platform is the inclusion of railway and tram track surveys. This is critical for District 1, where tram infrastructure is deeply integrated into the urban environment. To ensure temporal consistency with the ZuriACT ground truth points, imagery from the 2022 survey cycle is utilized for this thesis.

3.3.3. Other Data Sources

In addition to the primary 2022 screenshots obtained via the Infra3D web interface, iNovitas AG provides a large-scale dataset of raw street-view image tiles captured in 2024. This dataset comprises many high-resolution images covering the entirety of District 1, as well as an additional 1.4 km^2 area extending into neighboring districts (iNovitas AG, 2026a). A map showing the extending area can be found in Figure 4. These images are utilized in the transferability analysis to evaluate model performance and validate the robustness of the automated pipeline in unseen environments. The imagery is collected by specialized survey vehicles and trams traversing the urban road and rail networks. Unlike the static screenshots used for training, these images are discrete ”frames” extracted from multi-directional video sequences captured from three distinct camera angles. This raw dataset contains significant temporal and spatial redundancy as the survey routes could be repetitive. These raw images require an extensive pre-processing workflow, detailed in the Transferability Analysis section (Section 3.7), to filter out unqualified frames and ensure a clean domain.

The transferability analysis requires ground truth crosswalk data to evaluate the pipeline. These data is retrieved from the municipal database of Stadt Zurich. This dataset contains geo-referenced point data of crosswalks in the entire city, but is not openly published yet.

Furthermore, supplementary geo-referenced data are available on Zurich Open Data Portal and municipal database of Stadt Zurich. These open-sourced datasets include city and district boundaries (GIS-Zentrum, Geomatik + Vermessung, Tiefbau- und Entsorgungsdepartement, 2026) and footpaths (Tiefbauamt, Kanton Zürich, 2026). These layers are utilized to generate visualizations for spatial interpretation.

3.4. Data Pre-processing

The data pre-processing stage transits raw geospatial data to model-ready inputs. The main procedure is to primarily refining the ZuriACT data to generate a structured crosswalk database, associating these spatial points with Infra3D street view imagery, and establishing the labels that serve as training and validating sets for the computer vision model.

3.4.1. ZuriACT Data Pre-processing

The point dataset of validated and clustered crosswalk and curb ramps are used to generate a comprehensive and consistent crosswalk dataset. The crosswalk dataset in ZuriACT contains most of the crosswalks within District 1, with a strong focus on crosswalks with zebra crossings. In this thesis, crosswalk is being defined as any pedestrian passage connecting two opposing curb ramps, allowing pedestrian to cross.

In complex urban environments crosswalks, crosswalk could consist multiple segments separated by refuge islands or tram stops. In such instances, each individual section is identified as a distinct crosswalk, as each represents a discrete connection between two curb ramps (see Figure 6). These crosswalks are considered as multi-sectional crosswalks.

During this phase, it is observed that some connecting curb ramps are not identified as crosswalks in the ZuriACT data (see Figure 5). Using QGIS, a manual digitization process is conducted to fill these gaps. This results in a finalized dataset of 502 crosswalks within District 1. Each entry is assigned a unique Crosswalk ID, and where applicable, a Section ID to keep track of multi-section crosswalks (see Figure 6b and 6c). These digitized data points serve as the primary targets for the subsequent image association and model training phases.

3.4.2. Infra3D Street View Image Retrieval

Once the digitized crosswalk GeoJSON file is finalized, it is integrated into the Infra3D platform. Using the geo-referenced points as spatial reference, screenshots of each crosswalk were being manually captured (see Figure 7). These images are assigned unique Image IDs to link them directly to the crosswalk point data, ensuring that each visual representation remains geo-referenced according to the original GeoJSON.

(a) (b) (c)

Figure 6. Example of crosswalk segmentation: (a) a multi-sectional crosswalk (Crosswalk ID: c12), (b) section 1 of c12 (Crosswalk ID_Section ID: c12_1), and (c) section 2 of c12 (c12_2).

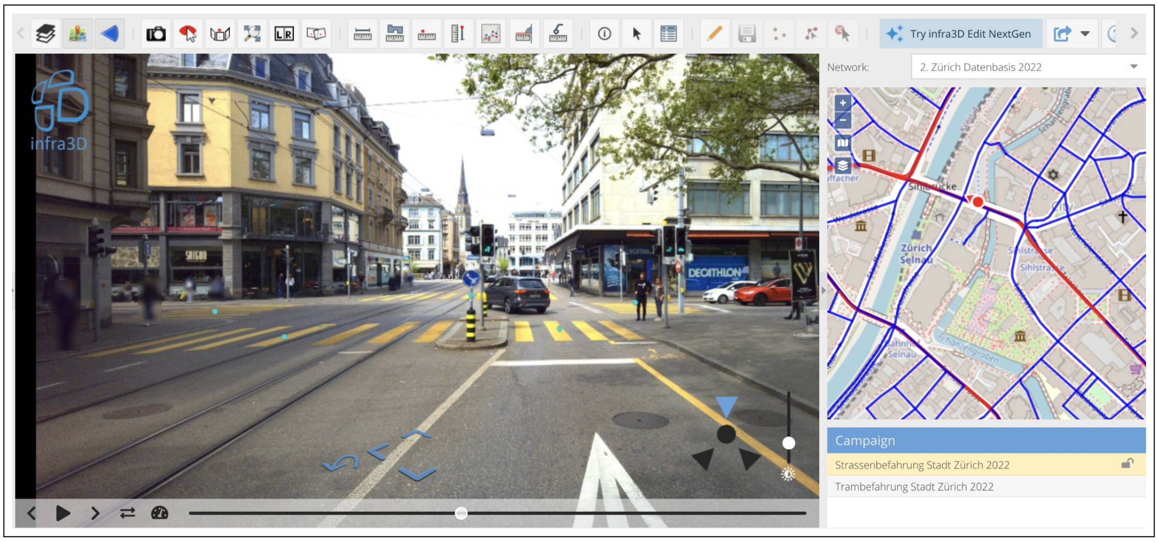


Figure 7. Interface of the Infra3D platform

3.4.3. Image Annotation and Data Preparation

With previous data pre-processing steps, a final dataset of 502 georeferenced images is uniquely paired with crosswalk points. To prepare these images for a YOLO (You Only Look Once) object detection architecture, manual annotation is done using a tool called LabelImg, a Python-based graphical annotation tool (Studio, 2015).

Four distinct classes are defined and manually annotated using bounding boxes.

- **Class 0 (Curb Ramp):** The ramps allowing pedestrians to enter the crosswalk.
- **Class 1 (Marked Crossing):** Specifically identifying the yellow ”zebra” stripes.
- **Class 2 (Tram Track):** Annotated as a single bounding box encompassing the track area, regardless of the number of lanes.
- **Class 3 (Traffic Light):** Signals relevant to pedestrian or vehicular movement at the crosswalk.

A sample of image with LabelImg interface is shown in Figure 8, where bounding boxes are being manually created closely around the features. The output of LabelImg consists of individual text files containing normalized pixel coordinates in the YOLO-specific .txt format.

For model development, the 502 annotated and georeferenced images are randomly distributed into a training set (80%) and a validation set (20%). Finally, a configuration YAML file is generated to define the data paths and class metadata. This concludes data pre-processing and the data preparation requirements for the object detection pipeline.

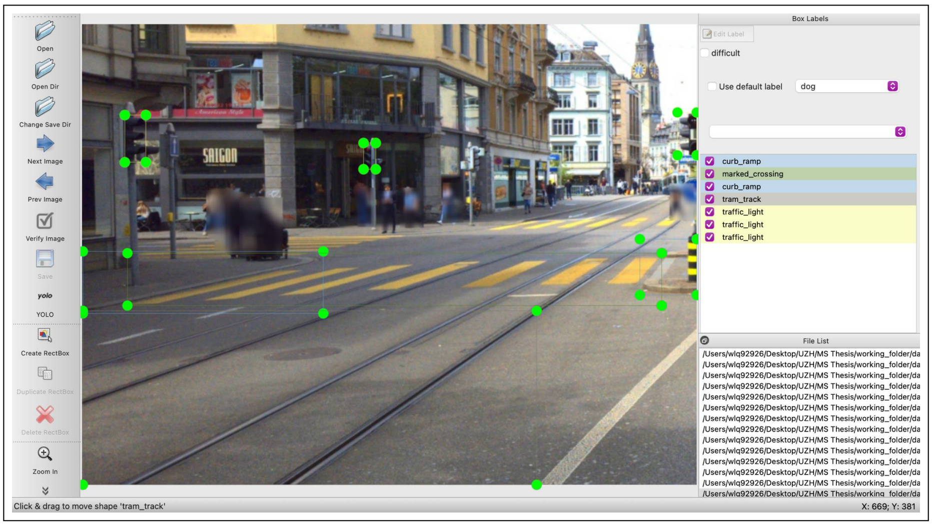


Figure 8. Interface of the LabelImg annotation tool

3.5. Computer Vision Model for Feature Detection

3.5.1. Model Configuration

Following the data preparation phase, The YOLO architecture is selected as the primary object detection framework. YOLO is a popular, high-speed, and accurate real-time object detection algorithm in computer vision, that identifies and locate objects in images in a single pass (Ultralytics, 2024). This thesis utilizes YOLO11, the latest iteration released by Ultralytics in 2024. YOLO11 introduces a more efficient backbone and neck architecture, optimizing feature extraction and gradient flow compared to its predecessors (Ultralytics, 2024).

During the experimental phase of configurating the model, both the YOLO11n (Nano) and YOLO11s(Small) variants are trained and evaluated. While the Nano version offers higher inference speeds, the YOLO11s model is selected for the final implementation. YOLO11s is chosen based on the size of the dataset and its priority in higher accuracy (Ultralytics, 2024).

The final hyperparameter configuration used for the baseline model is as follows:

- **Model:** YOLO11s (pre-trained on COCO)
- **Epochs:** 100
- **Batch:** 8
- **Learning rate:** 0.001

3.5.2. Model Augmentation

To enhance the model's ability to generalize and to prevent overfitting, several augmentation techniques are implemented during the training process. These methods artificially increase the diversity of the training set by applying stochastic transformations to the original images. The specific parameters for these augmentations are summarized in Table 2.

Definition of augmentation techniques implemented:

- **Horizontal Flip (flip lr):** Mirrors the image along the vertical axis.
- **Vertical Flip (flip ud):** Mirrors the image along the horizontal axis.
- **Translate:** Shifts the image horizontally or vertically by a fraction of the total

image size.

- **Scale:** Adjusts the size of the image by zooming in or out by a certain factor.
- **Degree:** Randomly rotates the image within a range of certain degrees.
- **Mosaic:** Combines four training images into one.
- **Mixup:** Overlays two images on top of one another with varying transparency.

Table 2. Data augmentation parameters

Augmentation	**Value**
Flip lr	0.5
Flip ud	0.0
Translate	0.1
Scale	0.5
Degrees	10
Mosaic	1.0
Mixup	0.1

These specific augmentations are selected to simulate real-world variations, also ensuring the data consistency with physical reality. Horizontal flipping and rotation allow the model to recognize feature regardless camera perspective and image tilting. The transition and scaling ensure the model can detects objects at vary distances and positions within the frame. Mosaic and Mixup are particularly beneficial as they create highly complex composite samples, enriching the relatively small dataset of 502 images. Mosaic and Mixup can improve the model's robustness against visual noise and partial occlusions.

The parameter vertical flip (flip ud) is strictly set to 0.0. This prevents the generation of physically impossible training samples, such as an upside-down traffic light. This ensures the model respects the natural gravitational orientation. These augmentation settings ensure that the resulting YOLO11s model is resistant to the perspective shifts and lighting variations found in the Infra3D imagery across both road and tram track perspectives.

3.6. Crosswalk Severity Index

The Crosswalk Severity Index (CSI) is a composite indicator developed to evaluate the complexity and relative safety of a crosswalk, when crossing by mobility-impaired pedestrians. The index is specifically designed to assess the challenges faced by mobility-impaired pedestrians in navigating complex a urban environment. The CSI integrates four key parameters: The presence of zebra crossing, traffic lights, tram tracks, and the physical length of the crosswalk.

A primary goal of this research is to automate the calculation of these parameters within a singular pipeline. While the presence of markings, traffic lights, and tram tracks can be identified directly via the previously trained YOLO11s model, determining the crosswalk length requires an additional depth estimation layer to translate 2D image pixels into metric distances.

3.6.1. Crosswalk Length Estimation using Depth Estimation Model

The length of a crosswalk is defined as the direct distance between its two connecting curb ramps. For calculating the CSI, this distance should be derived directly from imagery without relying on pre-existing spatial coordinates. However, calculating real-world measurements from monocular images presents several challenges:

- **Perspective Distortion:** In 2D images, the distance between pixels does not scale linearly with depth; objects further from the camera occupy less pixel space than those closer to the foreground.
- **Scale Inconsistency:** Because the training images are captured via manual screenshots from the Infra3D platform, the zoom level scales vary significantly across the dataset.
- **Unreliable Reference Objects:** While features like zebra stripes or traffic lights have standardized real-world dimensions, they cannot serve as reliable reference object for scale in this case. Due to the perspective and zoom limitations mentioned above, the pixel size of these reference objects varies too inconsistently between images to calibrate a universal pixel-to-meter ratio.

To overcome these challenges, this thesis shifts the focus from absolute geodesic measurement to relative length ranking. Instead of seeking meter precision, the objective is to distinguish and categorize crosswalks by their length, from small, medium, to large. This thesis integrates ZoeDepth, a state-of-the-art framework designed specifically for zero-shot metric depth estimation (Bhat et al., 2023), to calculate metric 3D distances between pairs of curb ramps. Unlike traditional models that either prioritize relative depth or focus on metric depth, ZoeDepth model combines both approaches (Bhat et al., 2023). The model utilizes a novel metric bins module, which employs a bin adjustment design to maintain metric scale across diverse domains (Bhat et al., 2023). The pipeline queries the ZoeDepth depth map to estimate the 3D separation between the centroids of the curb ramp bounding boxes. Even with limitations such as inconsistent camera perspectives and varying zoom levels in the Infra3D screenshots, ZoeDepth is expected to provide robust metric length estimates due to its specialized training for high generalization across unseen domains. This estimated length is then used to classify the crosswalk into one of three length classes (L): short, medium, or long.

To verify the reliability of the automated length estimation, a correlation analysis is done between the distances predicted by ZoeDepth and the geodesic distances derived from the ZuriACT spatial data. Two standard correlation coefficient is calculated:

- **Pearson Correlation (r):** This is a measure of the strength and direction of a linear relationship between two variables (Benesty et al., 2009). In this thesis, r measures how consistently the ZoeDepth estimated meters change in direct proportion to the ZuriACT measured meters. The value is between 0 and 1, a value of 1 would indicate a perfect linear match.

- **Spearman's Rank Correlation (ρ):** This is a nonparametric rank statistic used as a measure of the strength of a monotone association. It is particularly appropriate when data distributions make Pearson's coefficient potentially misleading. Unlike Pearson, Spearman's ρ does not require the assumption of a linear relationship or normally distributed data; instead, it assesses how well an arbitrary monotonic function can describe the relationship by focusing on the relative positions (ranks) of the data (Hauke and Kossowski, 2011; Sedgwick, 2014). The value is between 0 and 1, a value of 1 would indicate the ranks of both variables are identical.

The correlation leads to positive results. This confirms that the pipeline can reliably distinguish between short and long crossings. Based on this correlation and a manual qualitative review of the image dataset, the estimated distances are categorized into three severity classes. Thresholds are established to ensure the most dangerous "long" crossings were clearly isolated. The classification logic is defined as shown in Table 3. The length of each crosswalk that is successfully detected from the YOLO11s model is estimated, and a length class is assigned.

Table 3. Crosswalk length classification

Severity Class	**Description**	**Metric Range (d)**
Class 1	Short	$d < 11.55$ meters
Class 2	Medium	$11.55 \leq d \leq 16.47$ meters
Class 3	Long	$d > 16.47$ meters

3.6.2. Index Calculation

All four parameters are being prepared with previous models, the final step of this pipeline is to compute the Crosswalk Severity Index (CSI) for images that have two opposing curb ramps detected. This index serves as a quantitative measure of the navigational difficulty a crosswalk poses to mobility-impaired individuals. The CSI value S ranges from 0 (low complexity/easy) to 1 (high complexity/hard). The formula is defined as:

$$S = w_M(1 - M) + w_{TL}(1 - TL) + w_{TT}(TT) + w_L(L) \tag{3.1}$$

where:

- $M = 1$ if crosswalk markings present, else 0
- $TL = 1$ if traffic light present, else 0
- $TT = 1$ if any tram track present, else 0
- Length (L):
  - Class 1 (short): 0
  - Class 2 (medium): 0.5
  - Class 3 (long): 1
- w_M: weight of marked crossing (0.236)
- w_{TL}: weight of traffic light (0.217)
- w_{TT}: weight of tram track (0.281)
- w_L: weight of length (0.266)

The parameters of the Crosswalk Severity Index (CSI) are selected based on their specific impact on the safety and navigational ease of mobility-impaired pedestrians. Zebra crossings (marked crosswalk, M) and traffic lights (TL) are classified as beneficial accessibility features that lower crossing complexity. By prioritizing the pedestrian's right of way, these elements facilitate a more efficient and secure crossing process. To reflect the fact that the absence of these safety features increases the overall severity of a crossing, the formula utilizes the inverse of these variables, $(1 - M)$ and $(1 - TL)$.

Conversely, the presence of tram tracks (TT) introduces physical obstacles that significantly increase crossing difficulty. Navigating tracks typically requires more time and can decrease the perceived and actual safety for those with limited mobility. Similarly, crosswalk length is a critical safety factor; shorter crossings are inherently more convenient and safer as they minimize the duration of exposure to vehicular traffic. Consequently, the length variable (L) is integrated into the formula as a positive contributor to severity, normalized according to the categorical classes defined in the previous section.

Each of these parameters is given a weight. The weight is calculated based on a perception survey conducted with 30 mobility-impaired pedestrians (Georgescu et al., 2026; Ballinari, 2025). This survey required participants to rate the perceived difficulty or barrier level of specific urban infrastructure conditions on a numerical scale. For barriers, the scale ranges from 1 (Fully passable) to 5 (Not passable). For facilitators, the scale ranges from 1 (Very facilitating) to 5 (Not facilitating) (Georgescu et al., 2026; Ballinari, 2025). The arithmetic mean score is calculated for each category based on all 30 survey responses to determine the average perceived severity of each feature. To ensure the CSI score remains bounded between 0 and 1, these mean scores are normalized by dividing each factor's individual mean by the total sum of all means, ensuring the weights sum to exactly 1.0 (see Equation 3.2). The calculated weight of each parameter is shown in Table 4.

$$w_i = \frac{\mu_i}{\sum_{j=1}^{n} \mu_j} \tag{3.2}$$

Table 4. CSI parameters, perception categroy, and weights

Parameter	**Perception Category**	**Calculated Weight**
M	No zebra crosswalk	0.236
TL	No accessible pedestrian signals	0.217
TT	Tram tracks at crosswalk	0.281
L	Long crosswalk	0.266

The CSI is calculated for crosswalks that are successfully and accurately predicted from the YOLO11s model. A Crosswalk Severity Index (CSI) map in District 1 of City of Zurich is generated with QGIS for visualization and further analysis.

The continuous severity index $S \in [0, 1]$ is normalized to a discrete five-point scale scale for further comparison analysis with the ZuriACT severity. The mapping is applied as shown in Equation 3.3. This linear transformation maps $S = 0$ to Level 1 (most accessible) and $S = 1$ to Level 5 (least accessible), with intermediate values distributed across Levels 2–4. The resulting integer scale is assigned qualitative labels: Level 1 (Very Easy), Level 2 (Easy), Level 3 (Medium), Level 4 (Hard), and Level 5 (Very Hard). Each bucket covers an equal interval of 0.25 in the original scale, making it a uniform linear discretizations (see Table 5).

$$\text{Level} = \text{round}(1 + S \times 4) \tag{3.3}$$

Table 5. CSI Normalization to Five-Point Scale

Level	**Range of S**	**Label**
1	$[0.000,\ 0.124]$	Very Easy
2	$[0.125,\ 0.374]$	Easy
3	$[0.375,\ 0.624]$	Medium
4	$[0.625,\ 0.874]$	Hard
5	$[0.875,\ 1.000]$	Very Hard

3.7. Transferability Analysis

Following the generation of the end-to-end automated pipeline for crosswalk detection and severity evaluation, a transferability analysis is conducted to assess the pipeline's robustness across unseen environments. This analysis utilizes the raw 2024 sequence imagery provided by iNovitas AG, consisting of high-resolution frames extracted from survey videos. Each image is associated with a unique combination of Camera ID, Street Map ID, and Stream ID. The survey infrastructure varies by vehicle type, with trams utilizing three cameras oriented toward the front-center, front-left, and front-right, while land cars are configured with front-center, back-left, and back-right camera orientations. The images are accompanied by JSON metadata files containing precise camera geospatial coordinates. These metadata files are used to filter images that are outside of District 1 for this transferability analysis. A spatial boundary map of the raw images is shown (Figure 4) and a summary is listed in Table 6.

The target testing domain covers 16 distinct street maps that fall either partially or entirely outside District 1. For this analysis, a specialized subset is selected. This subset comprises Street Map IDs 99, 101, and 102, focusing specifically on Camera IDs 7002 (front-left) and 7005 (front-center) mounted on tram vehicles. These selected images are processed through the trained YOLO11s model to detect pedestrian accessibility features. The resulting crosswalk identifications are spatially compared against the official crossing dataset provided by Stadt Zürich to validate the transferability of the model. Crosswalk Severity Index (CSI) is being calculated and normalized for the crosswalk detection results.

Table 6. iNovitas image dataset overview

Metric	**Count**	**Percentage**
Unique capture locations	179,236	—
Total images	316,606	100%
Inside District 1	147,220	46.5%
Outside District 1	169,386	53.5%

4. Results

4.1. Data Pre-processing

The spatial distribution and composition of the study dataset are illustrated in Figure 9. The visualization presents a side-by-side comparison of the 502 digitized crosswalk data points and the original ZuriACT crosswalk inventory. Within the digitized map, specific crosswalks are categorized by their Section ID, where blue markers indicate section 1, red markers indicate section 2, and green markers indicate section 3. Individual, standalone crosswalks are represented by black data points. The dataset contains 15 crosswalks consisting of three sections and 46 crosswalks composed of two sections. In this thesis, each individual section is considered as a distinct crosswalk entity.

A comparison between the two maps shows additional data points in the digitized version on the left. While the original ZuriACT crosswalk dataset includes 465 points, an additional 37 points are manually digitized by referencing ZuriACT curb ramp data. Furthermore, the map shows certain areas within District 1, specifically in the city's Old Town, have no crosswalk data points present. These areas are pedestrian-priority zones with an absence of survey data.

The 502 crosswalk dataset is loaded onto the Infra3D platform, where screenshots for 493 data points are successfully retrieved. These images are manually labeled using LabelImg. The label process results in a total of 1,770 object annotations distributed across four distinct classes, as summarized in Table 7. Within this thesis, a crosswalk is defined as the connection between two opposing curb ramps, leading to the annotation of 991 individual curb ramps, consisting of 492 pairs and 5 supplementary instances. Because only 492 images contains these essential paired curb ramps, this specific subset of 492 images is outputted for use in the model training phase.

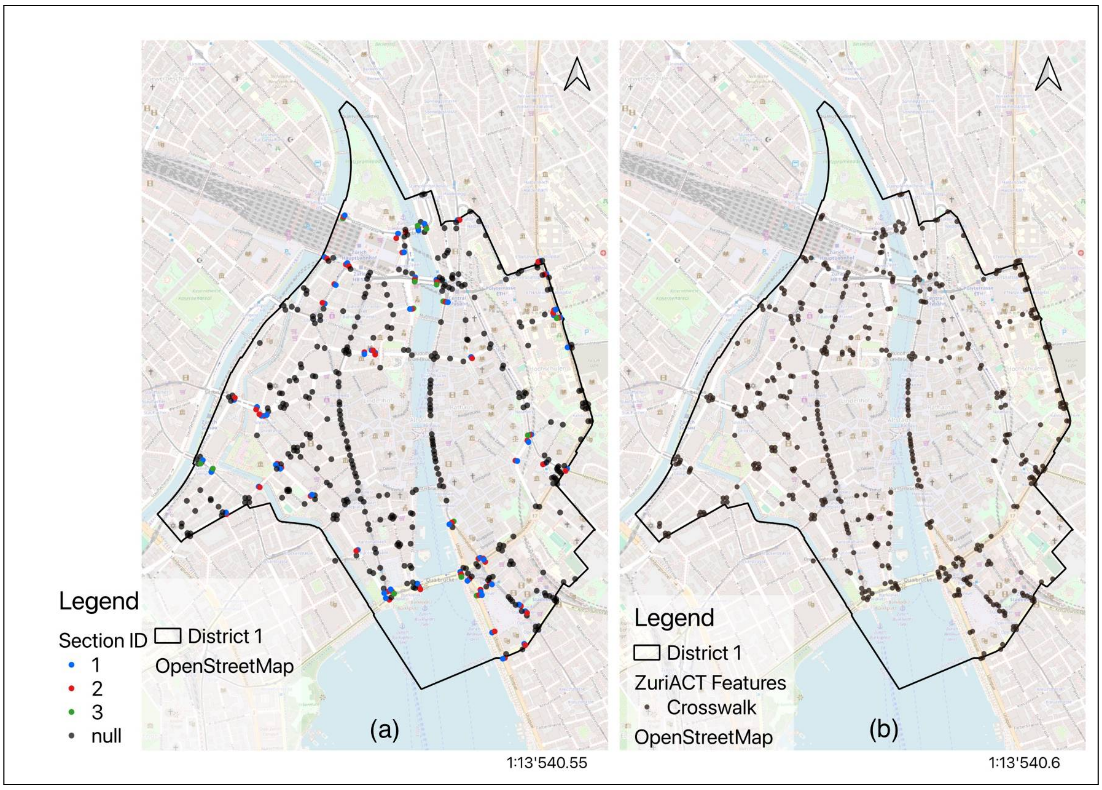


Figure 9. Comparison of a) digitized crosswalk dataset vs. b) ZuriACT crosswalk dataset

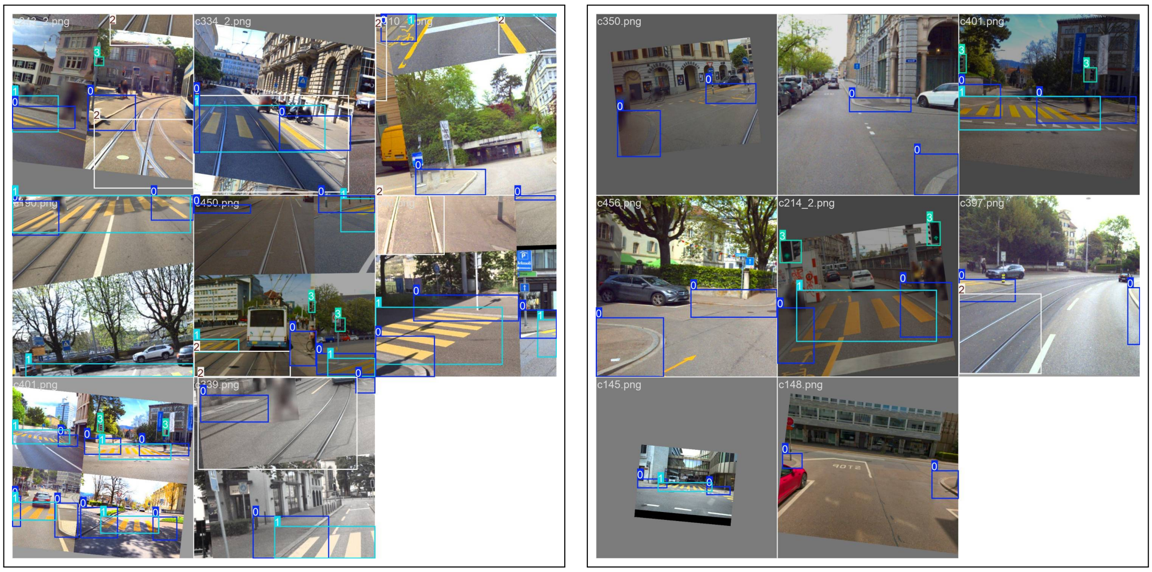
4.2. Model Evaluation

4.2.1. Model Augmentation

The learning process of the YOLO11s model illustrated augmentations. Figure 10 displays visualization of two training batches, batch 0 and batch 4412. The samples highlight the implementation of scale, translate, degree, and especially the mosaic data augmentation. The mosaic data augmentation technique combines four distinct training images into a single mosaic tile, which effectively forces the model to identify infrastructure objects across various scales and spatial contexts. Batch 0 (Figure 10a) represents a high-diversity batch where all four target classes are visible within multiple environmental settings. Figure 10b further demonstrates the model's exposure to challenging visual conditions, including variations in lighting, shadows, and perspective distortions. These augmentations are critical for developing a robust detection system capable of generalizing to the real-world complexities of Zurich's urban environment. By utilizing these augmented batches, the training process ensures that the model does not merely memorize specific street layouts. The model would learn the fundamental geometric and textural features of the pedestrian accessibility features.

Table 7. Manual annotation summary

Total images	**493**	
Class	**Annotations**	**Images**
Curb ramp	991	492
Marked crossing	311	309
Tram track	159	159
Traffic light	309	167



(a) (b)

Figure 10. Training data samples after augmentation: (a) batch 0, (b) batch 4412.

4.2.2. Model Performance

Figure 11 presents the training and validating curves of the YOLO11s model trained over 100 epochs across four object classes, curb ramp, marked crossing, tram track, and traffic light. All three training losses, including bounding box loss, classification loss, and distribution focal loss (DFL), exhibit a consistent and smooth downward trajectory throughout the training process. This indicates stable gradient updates and effective learning. The corresponding validation losses follow a similar pattern, decreasing sharply in the early epochs before plateauing approximately around epoch 40–50, with no subsequent upward trend observed. This behavior confirms that the model generalizes well to unseen data and shows no signs of overfitting.

In terms of detection performance, precision and recall both converge to approximately 0.85 and 0.88. Respectively, with some instability observed in the early

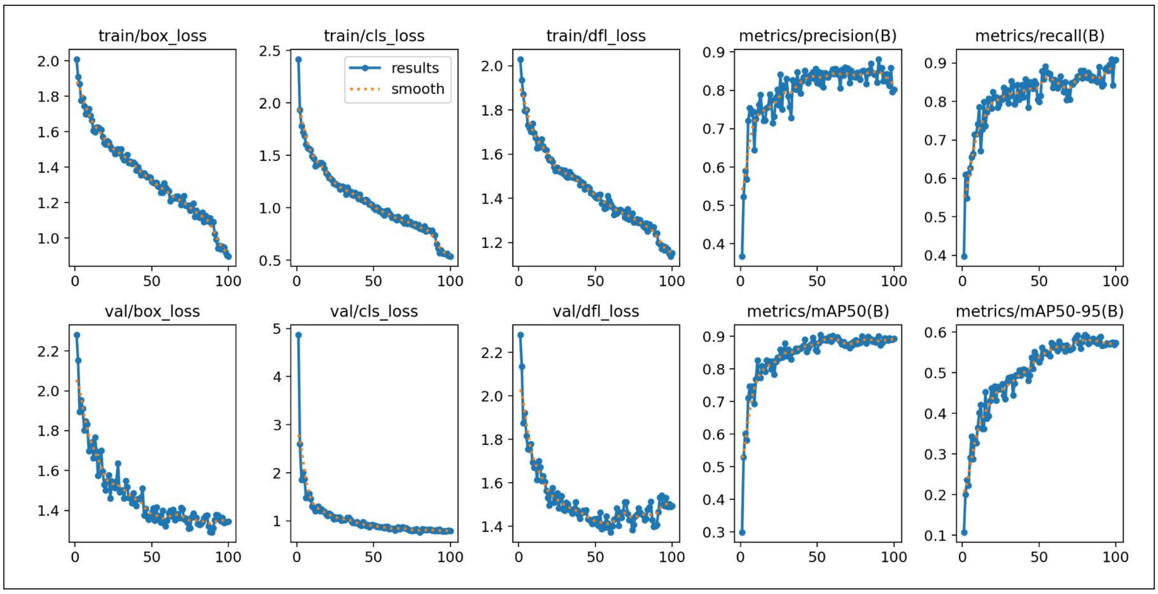


Figure 11. YOLO11s training and validation metrics

epochs. The mean average precision at IoU threshold 0.50 (mAP50) reaches approximately 0.90. The mAP50-95 converges to approximately 0.57. These results demonstrate that the YOLO11s model achieves strong detection performance despite its lightweight architecture. All metrics stabilizing in the later epochs, suggesting that convergence was reached within the allotted training duration.

Table 8 summarizes the detection performance of the YOLO11s model evaluated on the validation set for each class. The table reports four key metrics: precision (P), recall (R), mAP50, and mAP50-95. Precision (P) refers to the proportion of predicted bounding boxes that are correct, measuring the model's ability to avoid false positives. Recall measures the proportion of actual ground truth instances that were successfully detected, reflecting the model's sensitivity. The mean average precision at IoU 0.50 (mAP50) represents the area under the precision-recall curve at a fixed intersection-over-union threshold of 0.50. The mAP50-95 extends mAP50 by averaging mAP across IoU thresholds ranging from 0.50 to 0.95 in increments of 0.05. The mAP50-90 serves as a more rigorous measure of localization precision.

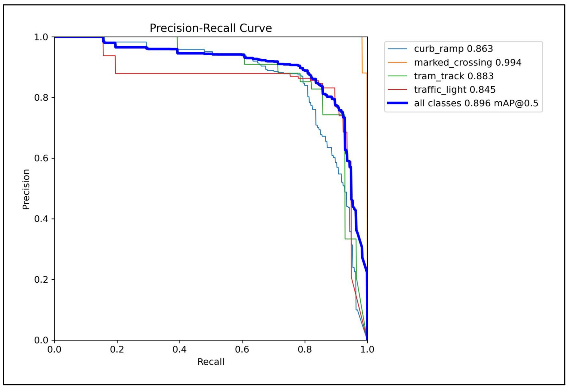
Across all 358 instances from 97 images, the model achieves an overall precision of 0.851, recall of 0.874, mAP50 of 0.896, and mAP50-95 of 0.593. At the class level, the marked crossing class yields the highest performance across all metrics, with a precision of 0.963, recall of 0.983, mAP50 of 0.994, and mAP50-95 of 0.791. The curb ramp class, despite having the largest number of instances (194), records a comparatively lower recall of 0.769 and mAP50-95 of 0.456. This suggested that the model struggles to locate curb ramps with high spatial precision. The tram track class achieves a precision of 0.733, the lowest among all classes. The traffic light

Table 8. Model performance metrics on validation dataset

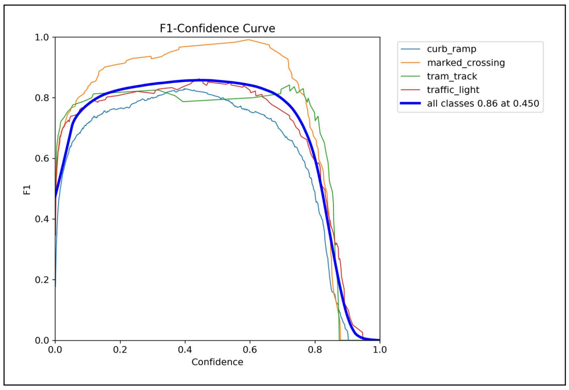
Class	Images	Instances	Box(P)	R	mAP50	mAP50-95
All	97	358	0.851	0.874	0.896	0.593
Curb ramp	97	194	0.877	0.769	0.863	0.456
Marked crossing	59	59	0.963	0.983	0.994	0.791
Tram track	28	28	0.733	0.857	0.883	0.552
Traffic light	41	77	0.830	0.888	0.845	0.575

class performs moderately, with a precision of 0.830, recall of 0.888, and mAP50 of 0.845, consistent with the overall model performance. Overall, these results indicate that while the model demonstrates strong average detection capability, performance varies across classes in a manner that reflects differences in instance count, visual complexity, and geometric regularity.

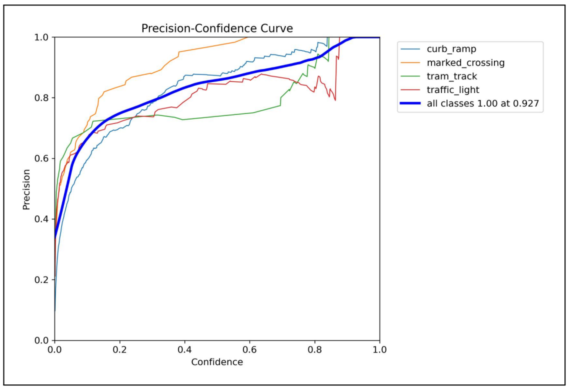
Figure 12 presents four diagnostic curves that further characterize the detection behavior of the YOLO11s model across varying confidence thresholds. The Precision-Recall (PR) curve in Figure 12a illustrates the trade-off between precision and recall for each class. A larger area under the curve indicates superior overall detection performance. The model achieves an overall mAP50 of 0.896, with the marked crossing class producing a near-perfect curve at 0.994, while curb ramp (0.863), tram track (0.883), and traffic light (0.845) exhibit steeper drops at higher recall values. The PR curves shows consistency with the per-class results (8). The F1-Confidence curve in Figure 12b plots the harmonic mean of precision and recall against the confidence threshold. This curve reveals that the optimal operating point for all classes combined is achieved at a confidence threshold of 0.450, yielding a peak F1-score of 0.86. The Precision-Confidence curve in Figure 12c shows that precision for all classes collectively reaches 1.00 at a confidence threshold of 0.927. This indicates that very high-confidence predictions are highly reliable, albeit at the cost of reduced coverage. Conversely, the Recall-Confidence curve in Figure 12d confirms that at a near-zero confidence threshold, the model achieves a recall of 0.96. The model captures the vast majority of ground truth instances, with recall declining gradually as the threshold increases. The curb ramp class exhibits the steepest recall decline among all classes, corroborating its comparatively lower recall observed in the quantitative evaluation (Table 8).



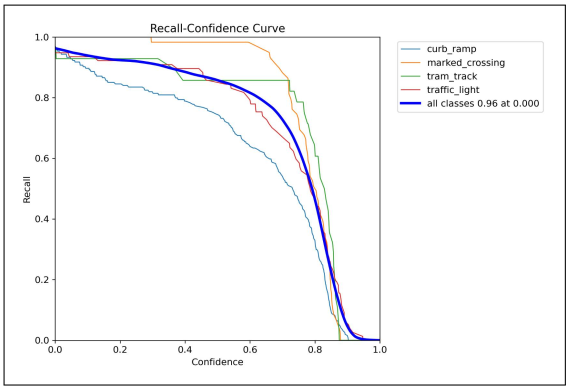
(a)



(b)



(c)



(d)

Figure 12. Performance matrics for the YOLO11s model with (a) Precision-Recall (PR) curve, (b) F1-Confidence curve, (c) Precision-Confidence curve, (d) Recall-Confidence curve.

The normalized confusion matrix is presented in Figure 13, providing further insight into the classification behavior of the model at the instance level. The diagonal entries representing correct predictions, are high across all four classes. The class of marked crossing achieves a perfect classification rate of 1.00, followed by tram track at 0.96, traffic light at 0.95, and curb ramp at 0.87. The most notable source of error is the false negative column corresponding to the background class. This indicates instances that are missed entirely by the detector. Curb ramp exhibits the highest miss rate, with 0.56 of its true instances being undetected and classified as background. This suggests that the model frequently fails to generate a detection proposal for curb ramps rather than misclassify it as another foreground category. Traffic light also shows a non-trivial background miss rate of 0.27, and respectively, tram track and marked crossing demonstrate comparatively lower false negative rates of 0.08 and 0.09. Inter-class confusion between foreground categories is minimal, with only a marginal misclassifications of curb ramp instances as tram track (0.04) and tram track instances as curb ramp (0.01). These findings suggest that the primary limitation of the model lies in its detection sensitivity rather than in its ability to discriminate between classes once a detection is made.

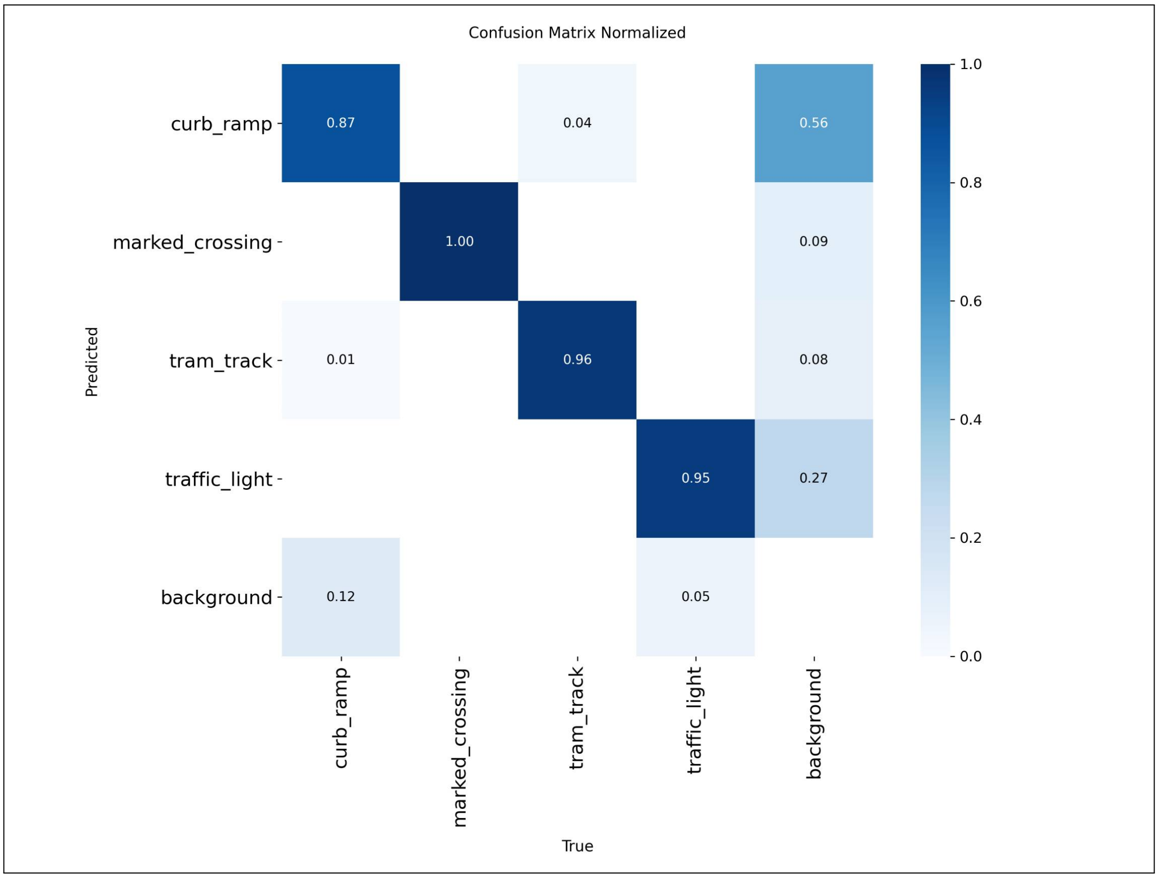
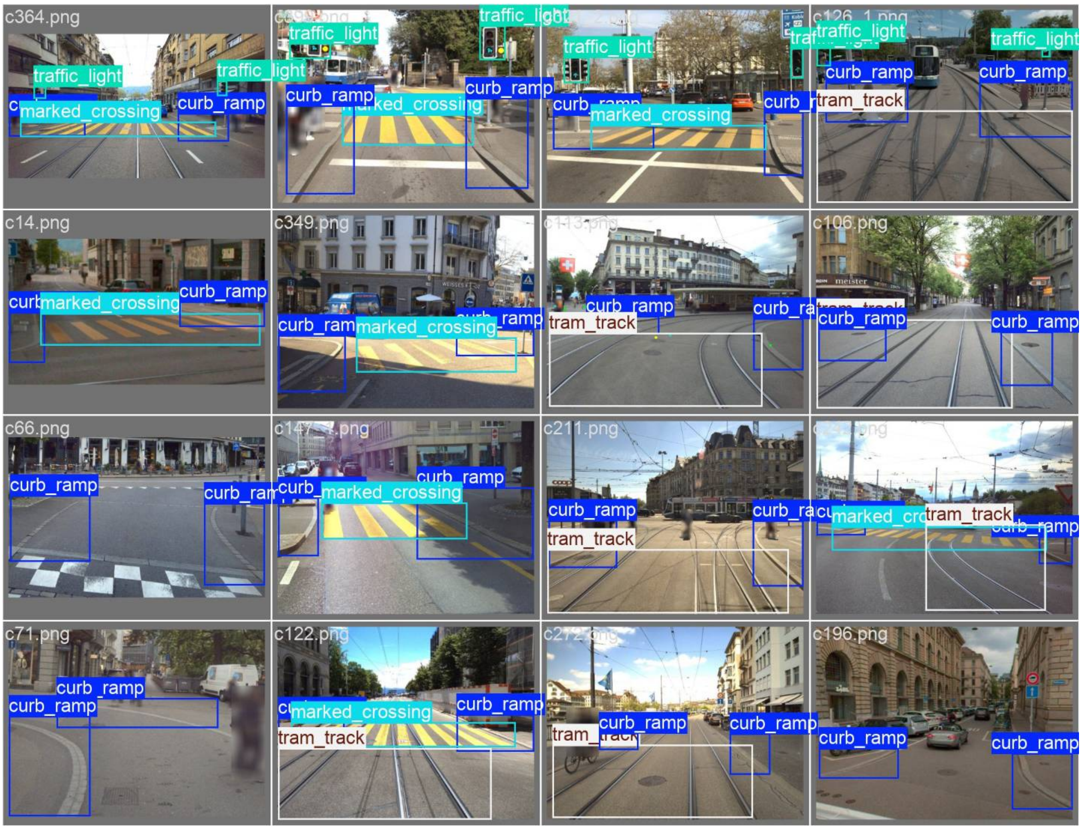


Figure 13. Normalized Confusion Matrix

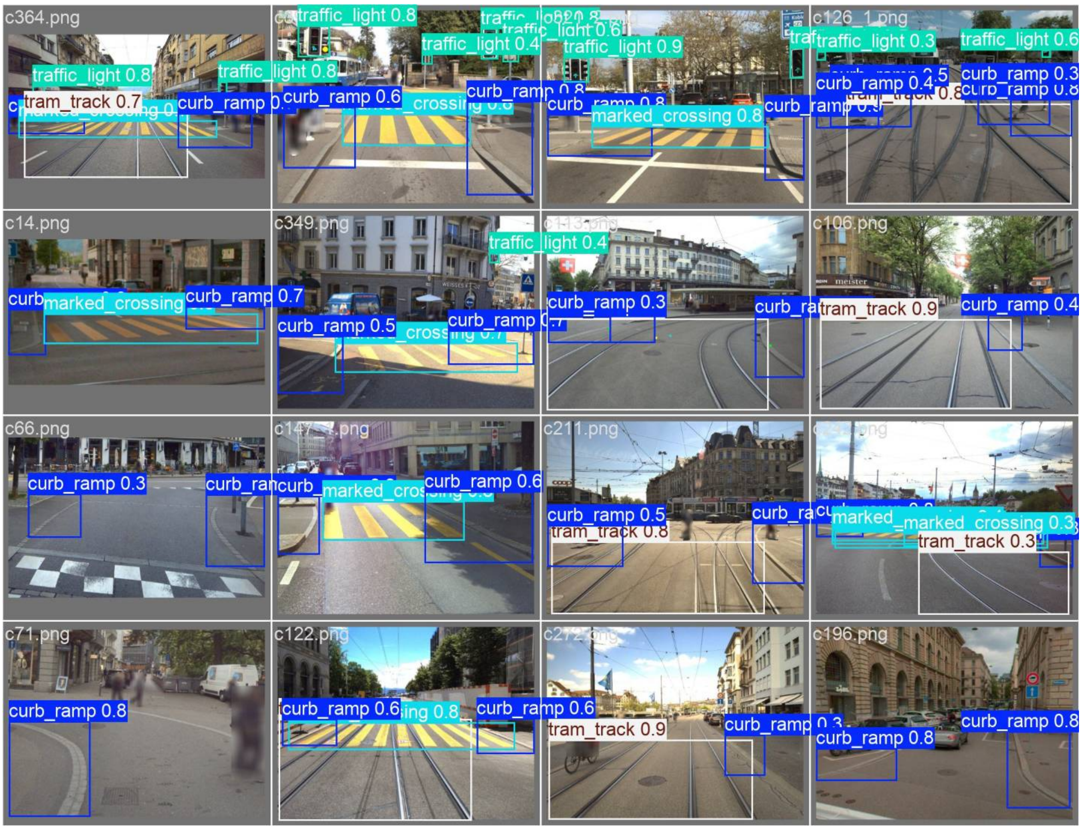
Figures 14 and 15 present qualitative comparisons between ground truth labels and model predictions for two validation batches. Figure 14a and Figure 15a indicate the labeled batch, and Figure 14b and Figure 15b indicate model predictions.

These visualizations provide a visual assessment of the YOLO11s model's detection behavior across diverse street scenes. In the majority of images across both batches, the predicted bounding boxes closely align with their corresponding ground truth annotations in terms of both spatial location and class assignment. This demonstrates that the model generalizes well to unseen scenes captured under varying lighting conditions, viewing angles, and urban environments. Visually distinctive classes such as marked crossing and tram track are consistently detected with well-fitted bounding boxes and high confidence scores, corroborating the strong quantitative performance (Table 8) reported for these classes. Curb ramps, which appear frequently across both batches due to their high instance count in the dataset, are generally detected correctly. However, some instances show lower confidence scores, particularly in scenes where the curb ramp is partially occluded, located at the image periphery, or visually similar to the surrounding pavement. Traffic lights are reliably detected in most scenes, with predictions closely matching the label positions.

Furthermore, a visual comparison between the label and prediction panels reveals that curb ramps and traffic lights exhibit a higher number of predicted instances than annotated ground truth boxes in several scenes. This observation indicates a tendency of the model to generate false positive detections for these classes, which is consistent with the normalized confusion matrix, showing that 12% and 5% of true background regions are respectively incorrectly classified as curb ramp and traffic light (see Figure 13). Overall, the qualitative results across both validation batches reinforce the quantitative findings, confirming that the model produces spatially accurate and semantically correct predictions under urban pedestrian environment.



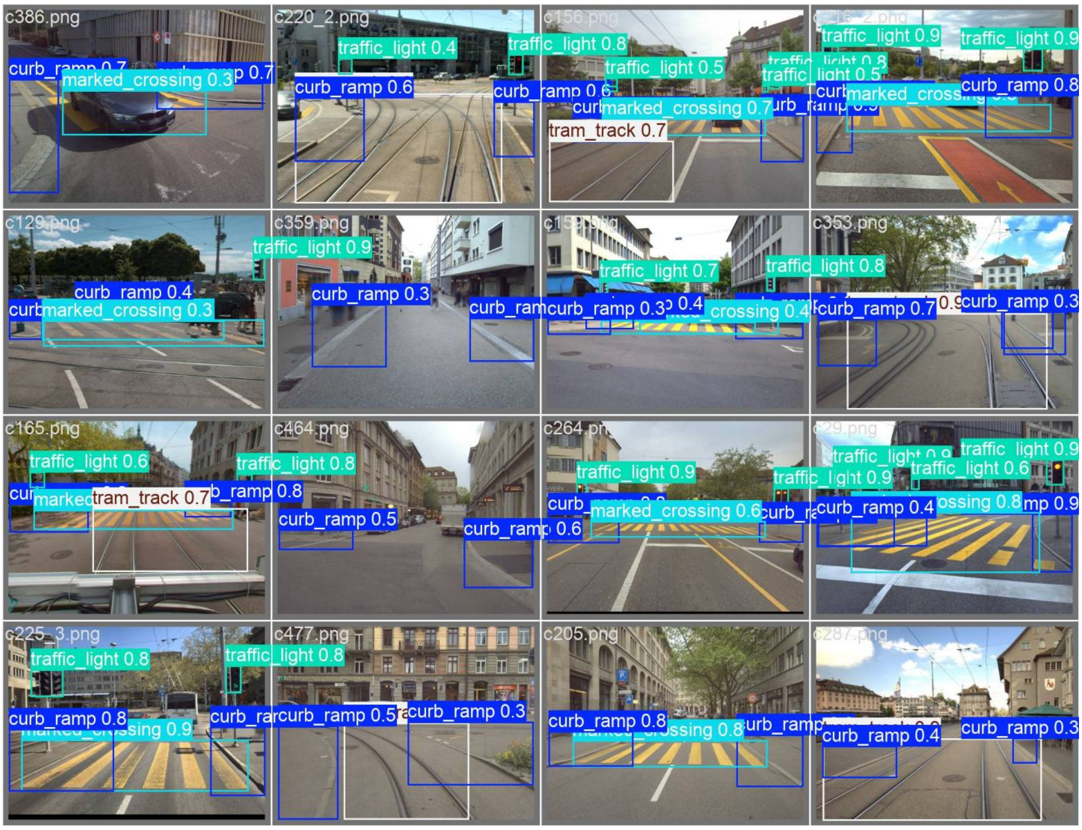
(a)



(b)

Figure 14. Validation batch 0, (a) labels vs. (b) predictions.

(a)



(b)

Figure 15. Validation batch 1, (a) labels vs. (b) predictions.

4.3. Crosswalk Severity Index

4.3.1. Crosswalk Length Estimation

To evaluate the reliability of the ZoeDepth-based depth estimation pipeline, the pseudo-distances derived from the monocular depth model were correlated against ground truth crosswalk lengths obtained from the ZuriACT dataset. Figure 16 presents the scatter plot of estimated pseudo-distance against ground truth crosswalk length, and Table 9 reports the corresponding correlation coefficients. A moderate positive correlation is observed, with a Pearson correlation coefficient of r $=$ 0.450 and a Spearman rank correlation coefficient of $\rho =$ 0.561, both accompanied by highly significant p-values of 4.861×10^{-23} and 1.674×10^{-37}, respectively. The higher Spearman coefficient relative to the Pearson coefficient suggests that the relationship between estimated and ground truth distances is better described as monotonic rather than strictly linear.

While the scatter plot reveals a generally increasing trend in estimated pseudo-distance with increasing ground truth crosswalk length, considerable variance is visible, particularly for ground truth values below 15 meters where the point cloud is densely concentrated. This indicates that the model captures relative depth ordering reasonably well but lacks the metric accuracy required for precise absolute distance estimation. Overall, the correlation results demonstrate that the ZoeDepth-based pseudo distances exhibit a statistically significant positive relationship with ground truth crosswalk lengths. This indicates that the rank ordering of estimated distances is broadly consistent with that of the true physical measurements. This property is sufficient to support the subsequent application of pseudo-distances as a relative proximity measure in the computation of the crosswalk severity index. In this case, the ordering and relative magnitude of distances are more critical than absolute metric precision.

Table 9. Correlation analysis results

Metric	**Value**	p**-value**
Pearson r	0.450	4.861×10^{-23}
Spearman ρ	0.561	1.674×10^{-37}

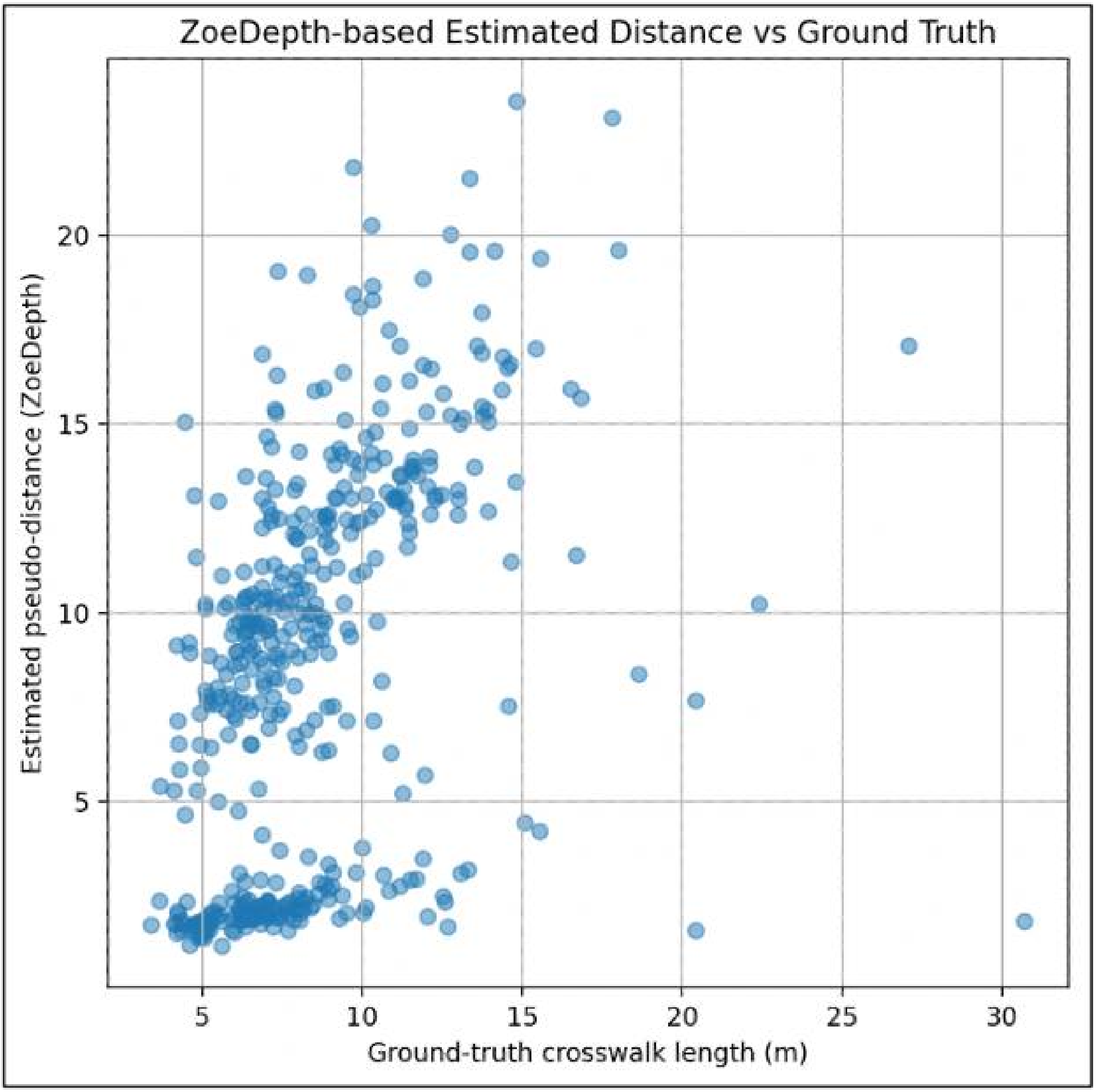


Figure 16. Correlation chart: pseudo distances vs. ZuriACT ground truth.

Following the length estimation and manual verification of the model detection outputs, crosswalk length categories are determined for all successfully detected instances. A total of 446 crosswalks across 446 images are accurately detected by the model and subsequently classified into three length categories: short, medium, and long. The distribution of crosswalk lengths is summarized in Table 10. Of the 446 detected crosswalks, 302 are classified as short (d $<$ 11.55 m), 105 as medium ($11.55 \leq d \leq 16.47$ m), and 39 as long (d $>$ 16.47 m).

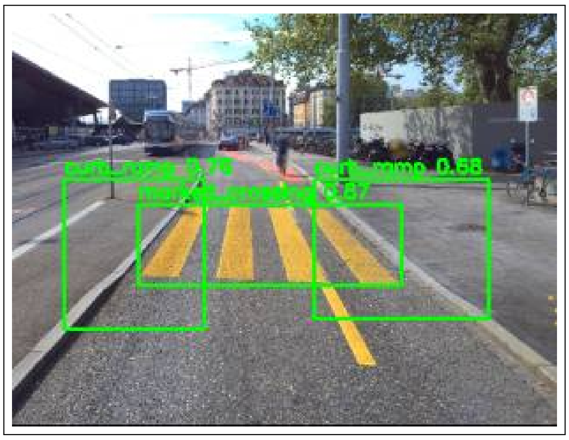
Representative sample images for each length category are illustrated in Figure 17. Short crosswalks generally correspond to single-lane road crossings, double tram track lanes, and narrower double traffic lane configurations (see Figure 17a and 17b). The medium crosswalks typically span double traffic lanes with an additional bicycle lane (see Figure 17c). Long crosswalks are associated with wider multi-lane road configurations (see Figure 17d). The predominance of short crosswalks in the dataset suggests that the majority of pedestrian crosswalk in the study area impose a relatively low crossing distance burden, which is friendly for mobility-impaired pedestrians. There are a total of 476 crosswalks being detected from the set of 493 images, but only 446 crosswalks have their length successfully classified. The rest of the images are being considered as medium for further severity index calculation.

Table 10. Distribution of crosswalk lengths

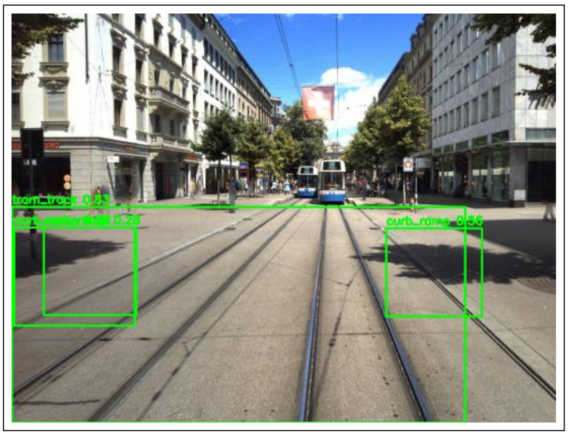
Length Category	**Number of Images**
Short ($d < 11.55$ m)	302
Medium ($11.55 \leq d \leq 16.47$ m)	105
Long ($d > 16.47$ m)	39
Total	**446**

4.3.2. Crosswalk Severity Index Visualization

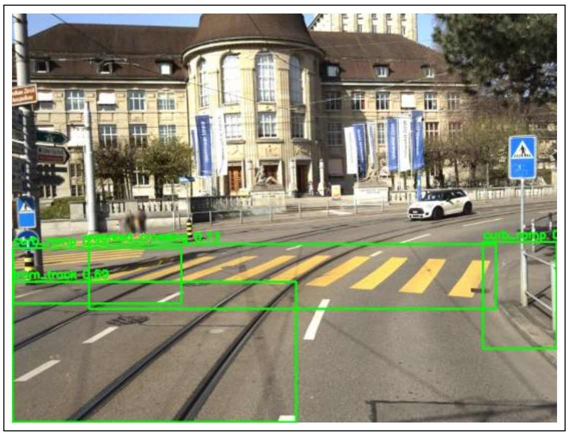
With the length categories derived from the depth estimation pipeline and the accessibility features detected by the YOLO11s model, the Crosswalk Severity Index (CSI) is calculated for a total of 476 crosswalks. The spatial distribution of CSI values is visualized in the map shown in Figure 18. where the color scale ranges from white (CSI = 0, lowest severity) to dark red (CSI = 1, highest severity). The spatial distribution of CSI values reveals a degree of geographic clustering, with lower severity scores concentrated in certain areas and higher severity scores appearing more continuously along specific corridors, as visible in Figure 18.



(a) Short



(b) Short



(c) Medium

(d) Long

Figure 17. Examples of crosswalk length categories: (a) short crosswalk, (b) short crosswalk, (c) medium crosswalk, and (d) long crosswalk.

The continuous CSI scores are subsequently normalized to a discrete five-point scale to enable direct comparison with the ZuriACT crosswalk severity ratings. The side-by-side spatial comparison between the normalized CSI and the ZuriACT severity is presented in Figure 19. Both maps share an identical legend ranging from severity level 1 (low complexity/easy, white) to severity level 5 (high complexity/hard, dark red). The color ramp grows gradually from white to dark red.

Table 11 summarizes the distribution of severity levels across both rating systems. The normalized CSI assigns the majority of crosswalks to severity level 2 (35.3%), followed by level 3 (25.4%) and level 4 (22.3%), with relatively few crosswalks at level 1 accounting for 16.2% and level 5 for only 0.8%. In contrast, the ZuriACT severity ratings are more heavily concentrated at level 1 (49.0%) and level 3 (33.5%), with considerably fewer crosswalks rated at levels 4 and 5 (3.0% and 0.2%, respectively). The two distributions differ notably in their central tendency, where the normalized CSI skews toward moderate severity levels 2 and 3, and ZuriACT assigns a near-majority of crosswalks to the lowest severity level. Despite these distributional differences, both systems assign very few crosswalks to the highest severity level 5, suggesting broad agreement on which crosswalks represent the most challenging cases.

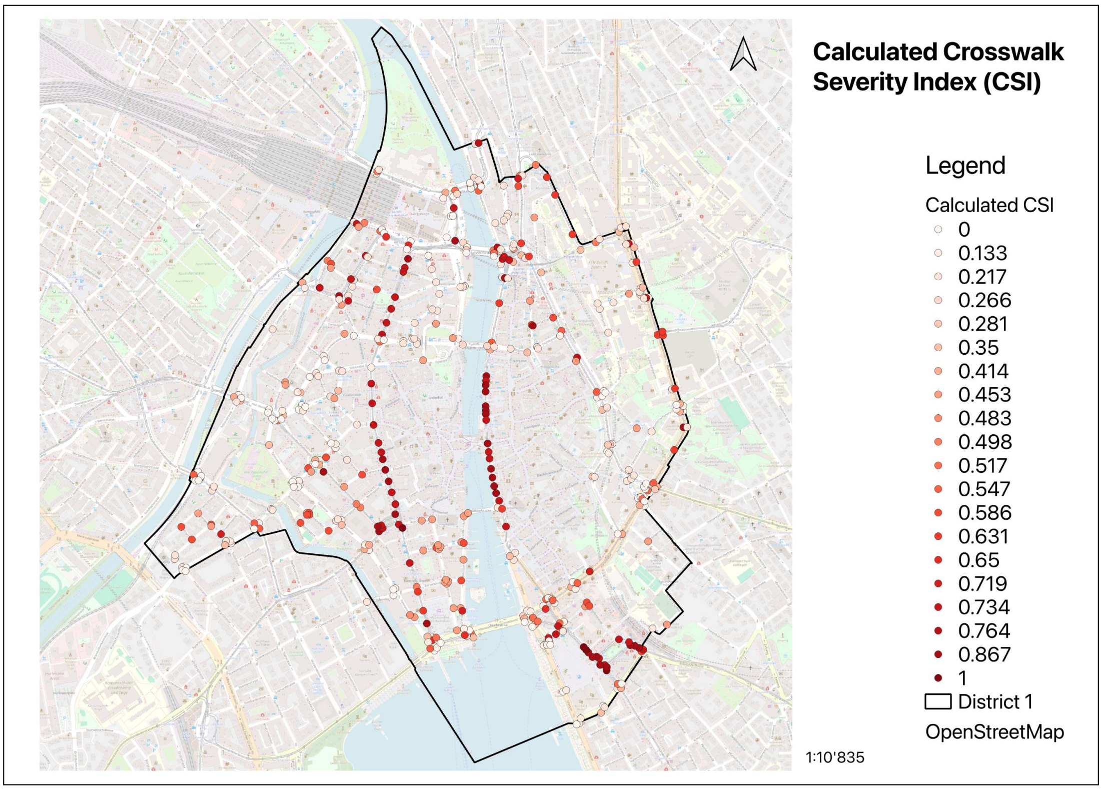


Figure 18. Calculated Crosswalk Severity Index: 0 (low complexity/easy) to 1 (high complexity/hard).

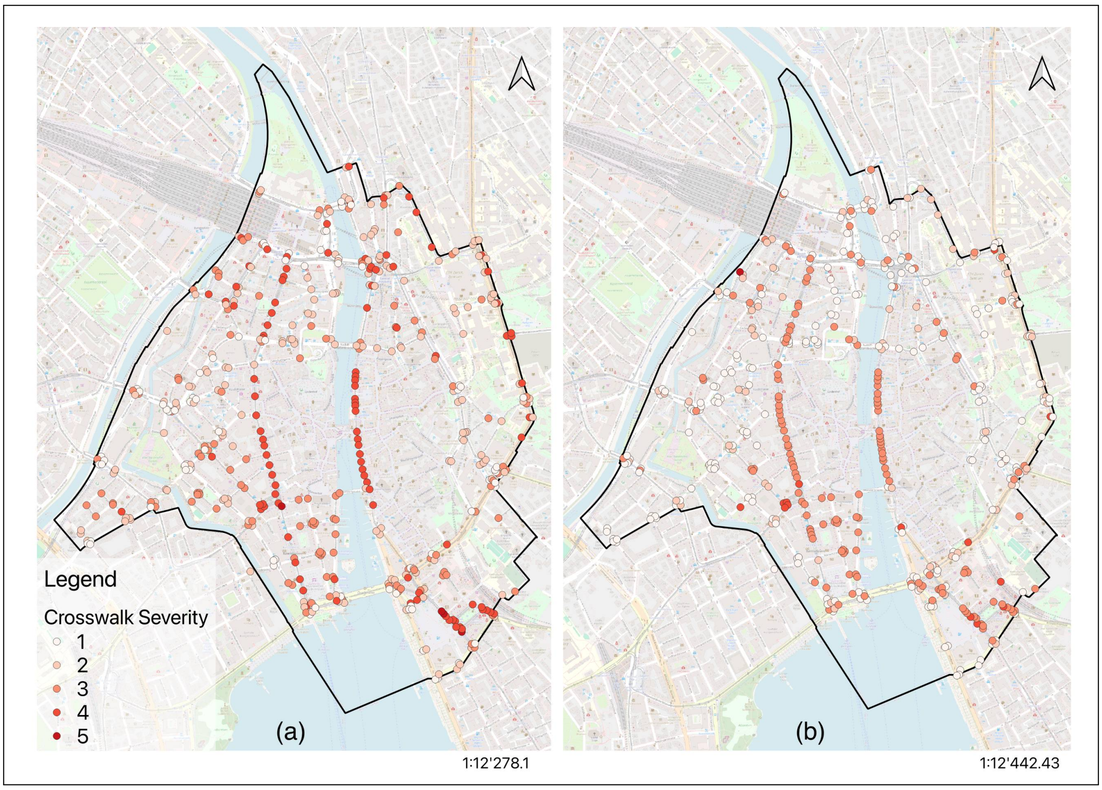


Figure 19. a) Normalized CSI vs. b) ZuriACT crosswalk severity: 1 as low complexity/easy, 5 as complexity/hard.

Table 11. Comparison of normalized CSI and ZuriACT severity

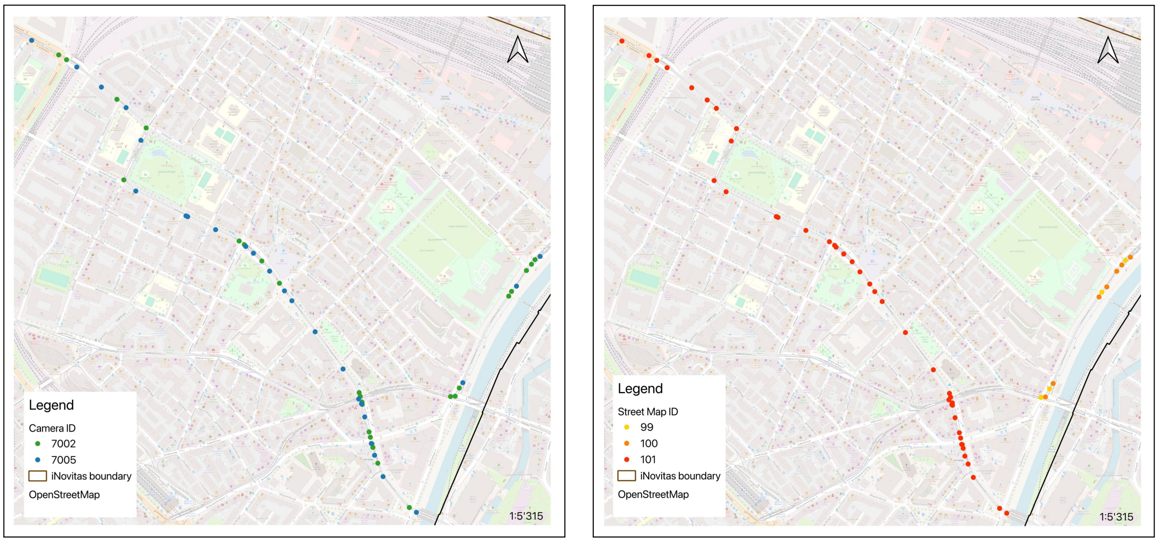
Severity Level	Normalized CSI Count	Normalized CSI %	ZuriACT Count	ZuriACT %
1	77	16.2%	228	49.0%
2	168	35.3%	66	14.2%
3	121	25.4%	156	33.5%
4	106	22.3%	14	3.0%
5	4	0.8%	1	0.2%
Total	**476**	—	**465**	—

4.4. Transferability Analysis

To assess the transferability of the developed pipeline, raw ZuriACT street view images from 2024 are filtered and utilized as input data. Approximately 16,120 images from Street Map IDs 99, 100, and 101, captured using Camera IDs 7002 and 7005, are processed through the YOLO11s detection model. Following filtering and accessibility feature detection, 58 crosswalk images are successfully extracted. The spatial distribution of detected instances by Camera ID and Street Map ID is visualized in Figures 20a and 20b, respectively.

Figure 21b provides a spatial overview of the detected crosswalk locations. For reference and validation purposes, the open crosswalk dataset published by Stadt Zürich (City of Zurich) is incorporated into the comparison analysis, as shown Figure 21a. The red dots represent crosswalks detected by the pipeline and dark gray dots indicate the Stadt Zürich reference crossings. A zoomed-in view of a representative intersection is provided in Figure 21c, enabling more detailed spatial observation and qualitative assessment of detection accuracy at the local level.

The 58 detected crosswalk instances are subsequently processed through the ZoeDepth model to estimate pseudo-distances and compute the corresponding CSI values. Around 48 pairs of opposing curb ramps are extracted from these 58 images, and CSI of each crosswalk is being calculated. The spatial distribution of the CSI is visualized in Figure 22a, with the normalized five-point severity ratings presented in Figure 22b.



(a) (b)

Figure 20. Detected crosswalks displayed in their (a) Camera ID, and (b) Street Map ID.

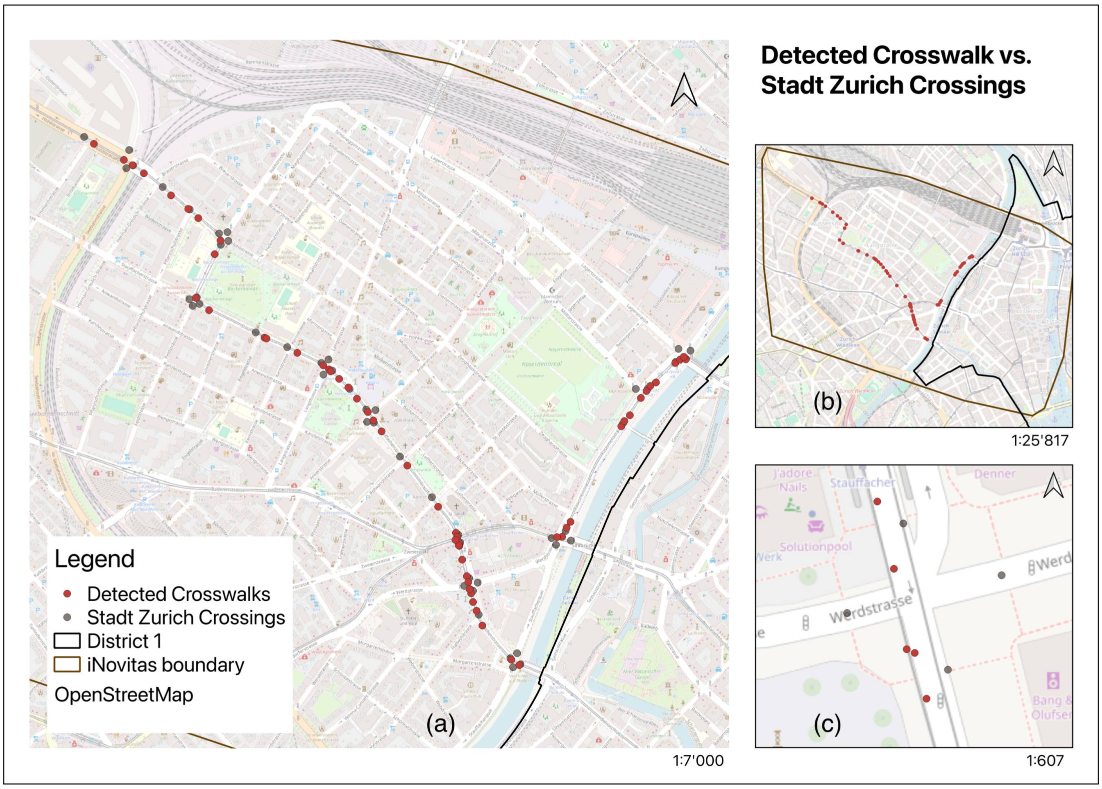


Figure 21. a) Detected crosswalks overlay with crossing data retrieved from Stadt Zurich, b) overview, and c) detailed view of one intersection.

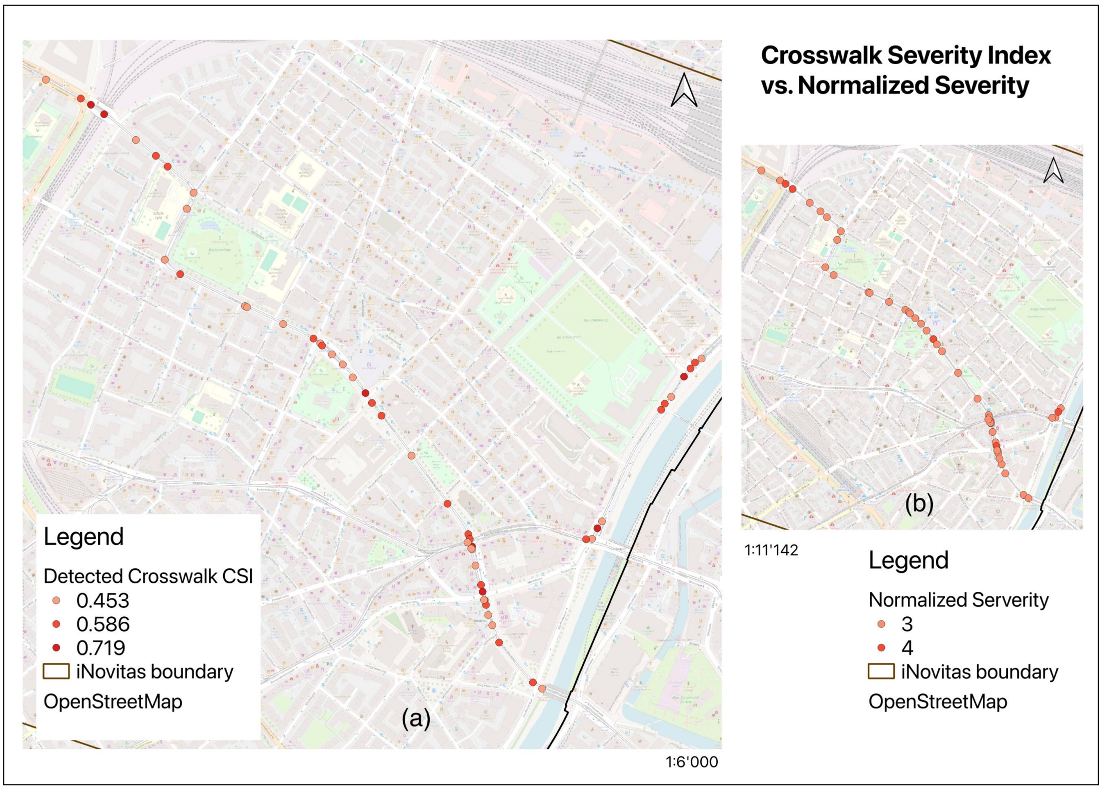


Figure 22. a) CSI calculated for the detected crosswalk, and b) normalized CSI.

5. Discussion

5.1. Result Analysis

5.1.1. Model Performance

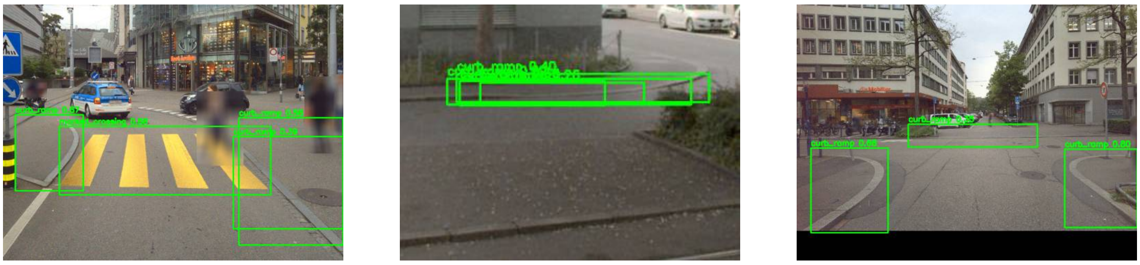
Detection Strengths. The YOLO11s model demonstrates strong overall detection performance across the four target classes, curb ramp, marked crossing, tram track, and traffic light. The trained model achieves a mAP50 of 0.896 and mAP50-95 of 0.593. This is notable given the relatively lightweight architecture of the model and the modest size of the training dataset. The high performance on the marked crossing class, can be attributed to the visually distinctive and consistent appearance of zebra-striped crosswalk markings, which provide clear and unambiguous visual features for the model to learn from. The zebra-striped markings are usually yellow in Zurich, where the consistency also increase the performance.

The high performance of the marked crossing class aligns with findings from previous research. Previous studies have demonstrated strong detection performance for crosswalk markings using YOLO-based models, CNN-based architectures, and ResNet models. In these studies, crosswalks are broadly defined as marked crossings and intersections, which is consistent with the marked crossing class in this thesis. (Zhang et al., 2022) achieved an average mAP50 of 99.0% under complex scenarios using CDNet based on YOLOv5, while (Lu et al., 2024) reported an mAP50 of 93.3 using X-CDNet based on YOLOX. The marked crossing class in this thesis achieves a mAP of 99.4% (see Table 8), exceeding the precision reported in these previous studies. This performance advantage is likely attributable to the homogeneous urban context of the study area, where all training and validation images were captured within the District 1 of Zurich. The relatively small validation set may also contribute to the observed high performance. Similarly, the strong performance on tram tracks and traffic lights reflects the relatively uniform visual characteristics of these classes across different scenes and lighting conditions.

Detection Challenges. The comparatively weaker performance on the curb ramp class, particularly its lower recall of 0.769, mAP50 of 0.863, and mAP50-95 of 0.456, warrants further discussion. Curb ramps exhibit considerable visual variability across the dataset, such as difference in surface texture, slope angle, and surrounding pavement context. These properties make it more difficult for the model to learn a robust and generalizable feature representation for this class. Furthermore, a high false negative rate is observed in the confusion matrix, where 56% of true curb ramp instances were missed and assigned to the background. This suggests that the model frequently fails to generate a detection proposal for this class rather than misclassify it as another class. This is a detection sensitivity issue rather than a classification one. Contributing factors include the small image area occupied by curb ramps, the low contrast between curb ramps and surrounding pavement, and the scale of the captured imagery relative to the camera position. Additionally, some lower curbs are incorrectly detected as curb ramps (see Figure 23c), likely representing vehicle access slopes rather than pedestrian crossing entrances, introducing false positives that further complicate evaluation.

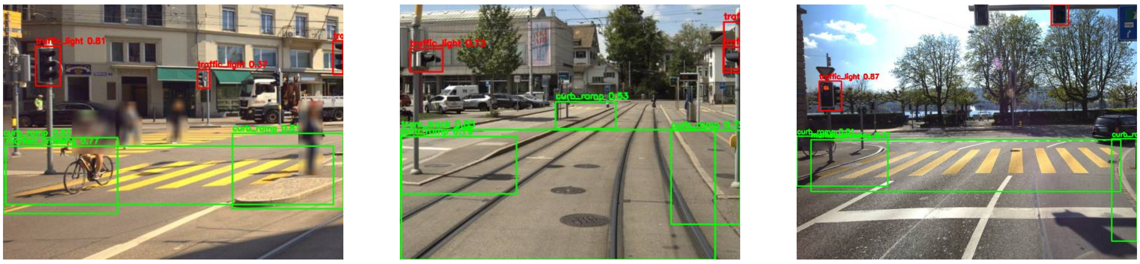
The challenge of curb ramp detection is well documented and discussed in the existing literature. O'Meara et al. (2025) trains a curb ramp detection model (RampNet, modified ConvNeXt V2), for detecting curb ramps over GSV panoramas. The study reports an AP of 0.924 for curb ramp detection using a dedicated large-scale dataset of 1,000 panoramas, substantially outperforming the prior state-of-the-art of 0.380 AP (O'Meara et al., 2025; Weld et al., 2019). The performance gap between RampNet and the curb ramp results in this thesis is therefore not surprising, as RampNet is trained on a significantly larger and purpose-built dataset specifically curated for curb ramp detection. This contextualizes the current results and highlights that curb ramp detection is an inherently challenging task, where larger and more targeted training datasets play a critical role in achieving higher detection performance.

The model has a tendency to generate false positive detections for curb ramps and traffic lights, as evidenced by the 12% and 5% background misclassifications rates respectively. As shown in a instance summary table generated from the model results (see Table 12), the instances of curb ramp and traffic light are higher than expected. This suggests that the model occasionally responds to background regions that share partial visual similarity with these classes. For example, pavement edges or distant street furniture are being detected as curb ramp (see Figure 23a and 23b), and other signals or background object in the images are being detected as traffic lights. As shown in Figure 24c, in addition to the pedestrian signal, other traffic light like objects are also being detected. Similar to curb ramps, duplication is one of the main factor causing the increasing instances (example see Figure 24b). Another reason of having more instances of traffic lights is the unexpected traffic light detection (see 24a), where pedestrian traffic light belong to another crosswalk is detected. This is also due to the camera position and frame of the images. These challenges also raise some uncertainty in further analysis, as the traffic light of certain crosswalks may not be identified correctly. This behavior is consistent with the qualitative observations from the validation batch visualizations, where additional predicted bounding boxes without corresponding ground truth annotations were visible for these two classes.



(a) (b) (c)

Figure 23. Example of detection error in curb ramps: (a) duplication in curb ramps (Crosswalk ID: c39), (b) missing curb ramps (Crosswalk ID: c5), and (c) wrong detection (Crosswalk ID: c35).



(a) (b) (c)

Figure 24. Example of detection error in traffic lights: (a) unexpected traffic light detected (Crosswalk ID: c307_1), (b) duplication in traffic lights (Crosswalk ID: c311), and (c) wrong detection (Crosswalk ID: c330).

RQ1. Overall, the trained YOLO11s model performs well in detecting visually distinctive and consistent accessibility features. Performance on curb ramps remains lower compared to other classes, due to their visual variability and subtle appearance. Despite this, the strong detection results across all classes support the use of the YOLO11s model for the subsequent CSI computation. In response to RQ1 (see Section 1.3), the results demonstrate that deep learning-based object detection can effectively identify pedestrian accessibility features from street view imagery, achieving an overall mAP50 of 0.896. While performance varies across classes, the model produces reliable detections for the majority of accessibility features relevant to crosswalk assessment, confirming the suitability of YOLO11s as the detection backbone of the proposed pipeline.

Table 12. Manual annotation summary vs. model detection instance counts

Class	**Manual Annotations**		**Model Detections**	
	Annotations	**Images**	**Instances**	**Images**
Curb ramp	991	492	1050	486
Marked crossing	311	309	321	309
Tram track	159	159	178	170
Traffic light	309	167	365	183
Total images	**493**			

5.1.2. Crosswalk Severity Index

The following section discusses length estimation results, crosswalk length classification, CSI calculation results, and the subsequent comparison with existing perception-based severity ratings. Together, these results address RQ2, where the CSI reflects crosswalk severity for mobility-impaired pedestrians (see Section 1.3).

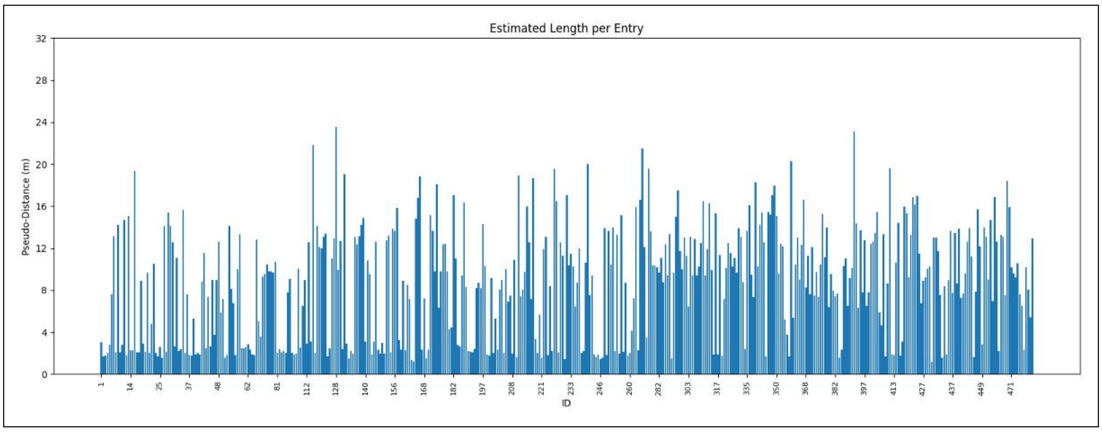
Length Estimation. The results of the depth estimation and length classification lead to further discussions. The implementation of the ZoeDepth model provides several benefits (Bhat et al., 2023), including combination of relative and metric depth estimation, allowing zero-shot generalization, and having high computational efficiency. Firstly, the ZoeDepth model combines relative and metric depth estimation, allowing for extraction of absolute distances in meters without requiring complex stereo-camera calibration. Secondly, ZoeDepth excels in ”zero-shot” scenarios, meaning it can accurately estimate depth in urban environments like Zurich even if it wasn’t specifically trained on those exact streets. In the situation of the thesis, no further training was needed to utilize the ZoeDepth model. Thirdly, the model also offers a high balance of speed and accuracy, approximately 0.17s per image, which is crucial for processing large-scale dataset (Bhat et al., 2023). Since the ZoeDepth is directly implemented without a training process, a statistical validation was required.

Length Classification. From the correlation analysis, with a Pearson r $=$ 0.450, and $\rho =$ 0.561, the analysis generally accept the ZoeDepth’s metric depth. The higher Spearman coefficient suggests the rank ordering is more consistent than the linear relationship, which is expected given the absence of metric camera calibration. As the severity index uses length as an ordinal classification rather than an absolute measure, the Spearman coefficient is considered the more appropriate metric, and the moderate rank agreement is deemed sufficient for this application. A Spearman correlation scatter plot is shown in Figure 16. The scatter plot reveals two distinct clusters in the estimated values, one dense group around 1-3 m and a larger group around 8-15 m. Based on the correlation results, a comparison between the pseudo-distance and the calculated ZuriACT distance is being conducted and generates length classes for further CSI calculation. Figure 25a illustrates a chart showing pseudo-length of each crosswalk, while Figure 25b shows the ground truth length calculated from the ZuriACT dataset. From the charts, it is observed that the pseudo-length has larger distinction between short crosswalk and long crosswalk. This can show that the pseudo-length may not be able to represent the actual length, but demonstrates a general ranking of the entire dataset. The larger distinction makes it easier to classify the pseudo-distances to small, medium, and large categories (see Table 3).

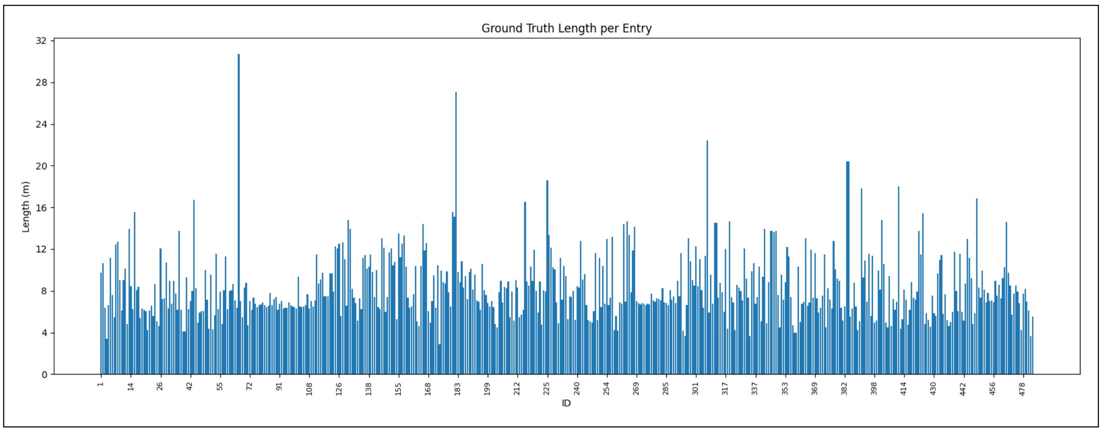
The distribution of length classes is shown in Table 10, having 302 short, 105 medium, and 39 long crosswalks. The results conclude that around 67% of the studied crosswalks within District 1 is considered as short crosswalk. The high number of short crosswalk is partially due to the definition of crosswalk in this thesis. This thesis consider each section of an entire crosswalk as a separate entry, crosswalks taken into consideration might not be the path crossing from one side of the road to another (see Figure 6). Cropping the crosswalks into sections highly increase the occurrences of short crosswalks within the entire dataset. This could be a major limitation, more details are being discussed in the following section (see Section 5.2) associate with possible solutions. Another reason would be that the study area (District 1) is the city center, including a significant part of the Old Town, leading less wide and multi-lane roads within the area.

The length of the crosswalk is highly related to the severity of crosswalk, and is used as a parameter in the CSI calculation. The longer crossing distances represent higher risk as they increase the ”conflict time” where a pedestrian is exposed to vehicular traffic. Conversely, short crosswalks reduce traffic exposure, while segmented crosswalks with pedestrian refuge islands provide additional reaction time for mobility-impaired pedestrians between crossing stages.

CSI Analysis. While the estimated length class is used as a parameter, the existence of marked crossing, traffic light, and tram track are also considered in the CSI. After calculating and normalizing the CSI, it is being spatially mapped. The spatial distribution of CSI values presented in Figure 18 reveals interpretable geographic patterns that reflect the underlying crosswalks of District 1. The map shows that crosswalks with low severity tend to be spatially co-located, especially at the intersections. The clustering of low severity crosswalks in certain zones suggests that pedestrian accessible intersections are well established, where accessibility features and facilitating infrastructure are existing at these intersections.



(a)



(b)

Figure 25. Comparison of crosswalk lengths: (a) ZoeDepth estimation, (b) ZuriACT measurement.

This is consistent with urban planning practices in which accessible intersections are often implemented systematically at the neighborhood or district level rather than in isolation (Boeing, 2021; Aghaabbasi et al., 2017). This is resulting in spatially contiguous zones of higher pedestrian accessibility, and suggesting crosswalks at intersections could be safer and more connivent for mobility-impaired users. Conversely, the spatial continuity of high severity crosswalks are along the major road corridors and primary tram routes. This indicates that these infrastructural axes present systematic accessibility challenges. This pattern is likely attributable to the wider carriageways, long coverage of tram tracks, and complex crossing context associated with main roads. From the map (Figure 18), two long roads with crosswalks in darker red could be observed. The concentration of tram infrastructure along these corridors further compounds accessibility difficulties, as tram tracks introduce additional surface irregularities that may pose navigational challenges for mobility-impaired pedestrians. Taken together, these spatial patterns suggest that the CSI is capable of capturing meaningful geographic variation in crosswalk accessibility and severity. The targeted infrastructure interventions along major road and tram corridors would yield the greatest improvements in pedestrian accessibility across the study area.

Normalized CSI Analysis. While the raw CSI map provides granular insight into the spatial distribution of severity across continuous values, the normalized CSI enables direct comparison with existing benchmark data. Figure 19 presents the two maps side by side, illustrating the normalized CSI and the severity ratings from the ZuriACT crosswalk dataset simultaneously. The ZuriACT severity levels are manually labeled and validated by mobility-impaired pedestrians, where level 1 denotes fully passable, level 2 almost fully passable, level 3 passable, level 4 almost not passable, and level 5 not passable, as defined in the labeling manual (Allahbakhshi, 2025; Georgescu et al., 2026). The two systems exhibit comparable distributional profiles (see Table 11).

The ZuriACT ratings show a pronounced concentration at the lower severity levels, suggesting a general tendency towards more favorable accessibility assessments. This bias is consistent with observations reported in human-centered accessibility surveys, where respondents may unconsciously anchor their ratings towards more moderate or positive evaluations (Schwarz, 1999), particularly when crossing conditions are perceived as adequate rather than severely problematic. In contrast, the normalized CSI distributes crosswalks more evenly across the intermediate severity levels, reflecting a more differentiated assessment of accessibility conditions. Despite these distributional differences, both systems converge in assigning a very small proportion of crosswalks to the highest severity level. This suggests broad agreement that severely inaccessible crosswalks constitute a small but non-negligible minority of the current pedestrian infrastructure. Both systems similarly indicate that the majority of signalized crossings in the study area are at least partially accessible to mobility-impaired pedestrians under current conditions.

It is worth noting that the CSI may offer a more comprehensive indicator of crosswalk accessibility than human-assigned severity ratings alone. The ZuriACT severity is derived exclusively from qualitative user perceptions. It could be inherently subject to individual variability, fatigue, and subjective interpretation of accessibility conditions. The CSI, by contrast, integrates quantitative geometric measurements derived from depth estimation, object detection outputs from the YOLO11s model, and qualitative perception-based survey data. The CSI is grounding the severity assessment in a combination of objective physical attributes and lived user experience, which adopt quantitative and qualitative metrics as similar approaches of previous studies (Ahmed et al., 2021; Kadali and Vedagiri, 2015). This multi-dimensional approach may account for the more distributed severity profile observed in the normalized CSI, and suggests that the CSI captures a broader range of accessibility dimensions than those reflected in perception-based survey ratings alone.

RQ2. Returning to RQ2 (see Section 1.3), the CSI demonstrates a capacity to reflect crosswalk severity in a qualitative, quantitative, and spatially meaningful way. The comparison with ZuriACT severity ratings shows broad agreement on the most critical crosswalk cases, while the more distributed severity profile of the CSI suggests it captures physical accessibility dimensions beyond what perception-based ratings alone convey. Furthermore, the CSI is achieved entirely through an automated pipeline without manual intervention. Overall, it provides a valid, objective, and scalable indicator of crosswalk severity for mobility-impaired pedestrians.

5.1.3. Transferability Analysis

Table 13. Transferability test results summary

Class	Annotations	Images
Curb ramp	155	58
Marked crossing	94	52
Traffic light	79	33
Tram track	47	44
Total images	**58**	

This section evaluates the transferability of the developed pipeline by applying it to an independent street view dataset captured under different conditions, directly addressing RQ3 (see Section 1.3).

Crosswalk Detection Analysis. The transferability analysis is conducted using a deliberately constrained subset of the iNovitas 2024 street view imagery. The study utilizes Street Map IDs 99, 100, and 101 to evaluate the generalization of the developed pipeline beyond the primary study dataset. Due to computational costs and redundancy in analyzing the entire coverages of the 16 street maps, only three are being analyzed as samples in this thesis. Street map 101 is the main study target, as street map 99 and 100 extend into the area of District 1 (see Figure 20b). A fundamental characteristic of the street map route structure is that each Street Map ID defines a fixed traversal route without random turning at intersections. This refers that the tram-mounted cameras capture the scene in a single direction of travel along each corridor (see Figure 20a). As a consequence, only crosswalks oriented along the direction of travel and visible from the front-facing cameras are consistently captured. The crosswalks at intersecting roads that branch off laterally remain largely undetected (refers to Figure 21c). This directional constraint inherently limits the spatial completeness of the detected crosswalk coverage, and a complete crosswalk detection is expected with the entire dataset. A more comprehensive transferability analysis will require the inclusion of all 16 Street Map IDs covering the entire study area, which include a wider range of routes and traversal directions. This is beyond the scope of the current analysis given the focus on demonstrating pipeline feasibility rather than exhaustive spatial coverage.

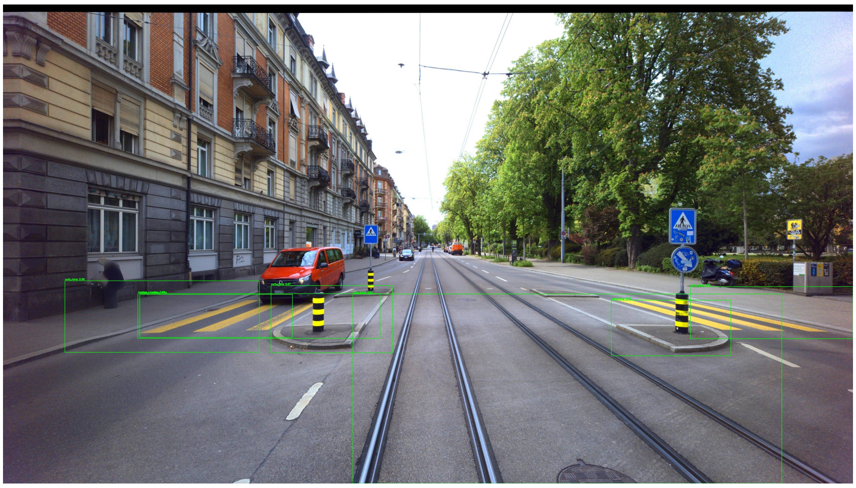
The spatial overlay of detected crosswalk locations against the Stadt Zürich crossing dataset reveals that the two point sets do not align precisely in geographic space (Figure 21a). The map shows that the detected crosswalk coordinates consistently offset from their corresponding reference positions. This discrepancy is primarily attributable to the nature of the coordinate extraction process, whereby the spatial location assigned to each detected crosswalk corresponds to the geographic position of the tram-mounted camera at the moment of image capture. The camera location is not the true ground-level location of the crosswalk itself. Since the crosswalk appears at a distance ahead of the camera in the captured frame, the recorded coordinate reflects the camera's position along the tram route rather than the physical location of the pedestrian crossing infrastructure. The magnitude of this spatial offset is therefore dependent on the distance between the camera and the detected crosswalk, which varies across instances. This introduces a systematic positional bias that is consistent in direction but variable in magnitude. The nature of having only camera location associate with the dataset making precise geo-referencing of individual crosswalk detections challenging. Additional depth-based coordinate re-projection, Geotagging, or GPS-anchored ground truth matching could possibly solve this problem (Krylov et al., 2018).

Dataset Advantages. The consistent mounting configuration of the tram-mounted cameras introduces both advantages and limitations in this transferability analysis. Camera IDs 7002 and 7005, corresponding to the front-left and front-center cameras mounted on trams, were selected for this analysis (refers to Figure 20a). They provide the most direct forward-facing view of the road ahead and are most likely to capture approaching crosswalks within the field of view. The fixed mounting height and orientation of these cameras on the tram ensures a degree of geometric consistency across all captured images. Unlike manually retrieved street view images used in the main study, the scale relationship between image pixels and real-world distances is relatively stable across the raw dataset. This consistency is beneficial for the ZoeDepth-based depth estimation component of the pipeline, as the model encounters a more uniform input distribution compared to hand-held or variable-height camera setups.

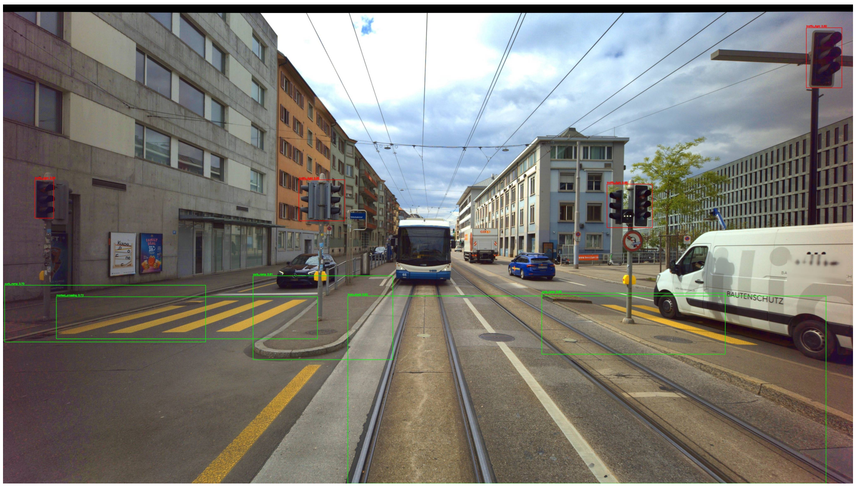
Dataset Limitations. However, the elevated and fixed mounting position of the tram cameras also introduces detection challenges for the YOLO11s model. Due to the consistent camera height and the relatively wide field of view, crosswalks and their associated accessibility features are often captured at a greater distance from the camera compared to images in the training dataset. This is causing these features to appear smaller and flatter within the image frame. This scale reduction is particularly problematic for curb ramps, which already occupy a small image area under normal conditions and may appear almost indistinguishable from the surrounding pavement surface in some of these images. This will highly increase the likelihood of miss detections. Additionally, the wider field of view afforded by the tram-mounted cameras means that a greater number of objects visually similar to traffic lights are possibly simultaneously visible within a single frame. Some examples are tram signal indicators, overhead line infrastructure, and street signage. It will potentially introducing false positive detections or causing the model to miss the intended targets.

These limitations are partly a consequence of the small subset of street map data utilized in this transferability analysis. By incorporating the full set of available street map routes and applying spatial de-duplication to merge detection instances within close proximity, a more spatially complete and accurate crosswalk dataset could potentially be constructed. Considering the entire dataset would mitigate some of the coverage gaps introduced by route directionality and scale variability. Nevertheless, the inherent geometric mismatch between the tram camera perspective and the camera configurations represented in the YOLO11s training data remains a concern that warrants further investigation. Two sets of sample image illustrating the scale and detection challenges described is provided in Figure 27 and 26.

Furthermore, because both Camera IDs 7002 and 7005 are mounted on the same tram and capture overlapping fields of view, the same physical crosswalk is frequently detected independently by both cameras as the tram approaches. This results in pairs of detection points in close spatial proximity to one another on the output map, representing the same crosswalk rather than two distinct instances. While this duplication does not affect the qualitative conclusions of the transferability analysis, it inflates the total count of detected instances and should be addressed through a post-processing de-duplication step in future implementations, for example by merging detections within a defined spatial proximity threshold. In this thesis, the post-processing of duplicate and redundant detections is performed manually. However, applying this process to the full raw dataset at scale would be considerably time-consuming. One practical solution is to sub-sample the image dataset prior to feeding it into the detection model. Since the raw images are extracted from continuous survey videos, the same crosswalk may appear across multiple consecutive frames. A sub-sampling strategy would be selecting the middle frame as the representative frame from every ten consecutive images. This could effectively reduce redundancy while retaining sufficient coverage for reliable detection. With an appropriate sub-sampling interval, crosswalk instances could ideally be detected without duplication, significantly reducing both computational cost and post-processing effort.

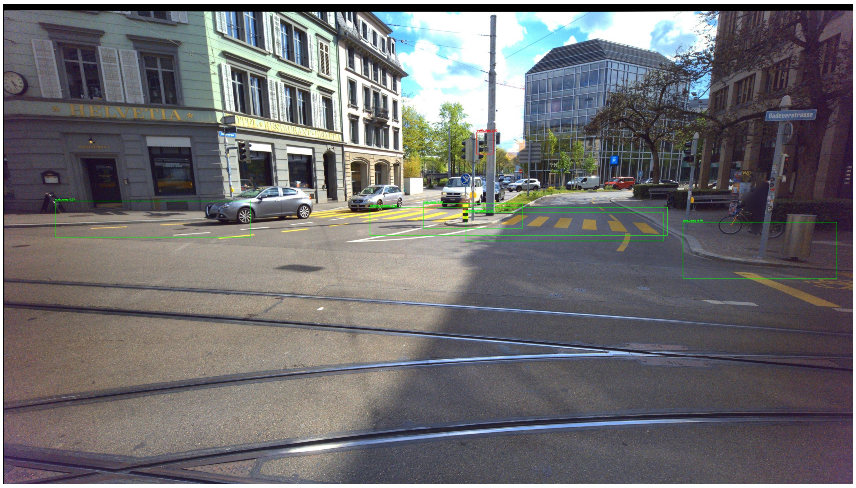


(a)



(b)

Figure 26. Examples of crosswalk detection from the front camera: (a) 101-7005-43851, (b) 101-7005-43360.



(a)

(b)

Figure 27. Examples of crosswalk detection from the front-left camera: (a) 99-7002-65112, (b) 101-7002-43529.

CSI Analysis. The CSI values computed for the crosswalks detected in the transferability analysis range from 0.453 to 0.719, corresponding to normalized CSI of 3 to 4 on the five-point scale (see Figure 22). This range indicates that the crosswalks detected along the evaluated tram corridors are predominantly of intermediate to moderately high severity. This suggests that while these crosswalks are generally passable, they present non-trivial accessibility challenges for mobility-impaired pedestrians. These findings are broadly consistent with the results obtained from the primary study area in District 1. The majority of crosswalks are also assigned to intermediate severity levels, with very few instances at either extreme of the severity scale. The convergence of severity profiles across the two datasets, despite differences in data collection methodology, camera configuration, and geographic coverage, provides preliminary evidence that the CSI pipeline produces consistent and contextually plausible severity assessments. This alignment supports the transferability of the developed methodology to other street view datasets captured under similar urban conditions, and suggests that the CSI is responsive to the physical attributes of the road environment.

RQ3. Overall, the transferability analysis provides general support for the reliability of the end-to-end pipeline, from YOLO11s-based feature detection through to CSI computation. The analysis supports the scalability and reproducibility of the proposed pipeline, and it is especially reliable when applied to street view imagery captured in a similar urban context. Returning to RQ3 (see Section 1.3), the developed automatic pipeline is transferable to independent street view datasets. While the current analysis demonstrates the feasibility of transferability, the robustness of the pipeline across a wider range of routes, camera configurations, and urban morphologies could be further established through additional evaluation on more extensive and geographically diverse datasets.

5.2. Limitations

While the the study is being successfully implemented with acceptable assessment result, there are some limitations remained.

Data Sources and Data Consistency. The imagery used in this thesis introduces several limitations related to data quality and consistency. The iNovitas Street View Imagery (SVI) is selected over Google Street View (GSV) to overcome known limitations of GSV, including incomplete spatial coverage and outdated imagery. However, as iNovitas imagery is not openly accessible, its usage remains a practical constraint for reproducibility and broader adoption of the pipeline. Since the ZuriACT dataset is collected based on SVI surveyed in 2022, screenshots from the infra3D platform are utilized to maintain data consistency, as raw iNovitas images from 2022 are unavailable and commercially costly to obtain. These screenshots introduce several limitations, including variable image scale, inconsistent camera positioning, absence of embedded geospatial coordinates, and potentially lower image resolution or motion blur. For the transferability analysis, raw images from 2024 are used, offering advantages in resolution and geometric consistency. However, processing the large volume of images generated from continuous surveying video incurs substantial computational and commercial costs. Some pre-processing or post-processing steps are needed, including frame extraction, filtering, de-duplication, and coordinate assignment. Sub-sampling as a optimal solution is discussed above (see Section 5.1.3). Reducing the study area is adopted as a practical solution in this thesis, which in turn introduces the spatial coverage limitations.

Targeted Data Augmentation. The visual variability of curb ramps presents a persistent detection challenge that could be partially addressed through targeted data augmentation strategies. These include the simulation of diverse surface textures, pavement materials, and a wider range of lighting conditions. Another augmentation is simple duplication of existing curb ramp instances and the incorporation of additional annotated datasets. As the current training dataset consist a relatively small range, the detection performance of the model is naturally constrained. Dedicated models trained specifically for curb ramp detection remain scarce in the existing literature. One of the few examples is RampNet, introduced by O'Meara et al. (2025), which utilizes GSV panoramas for large-scale curb ramp detection. However, RampNet is developed and evaluated in a North American urban context, and its transferability to European cities remains an open question. European cities such as Zurich have crosswalk infrastructure, pavement standards, and urban design conventions which differ considerably from North American urban context. The pipeline introduced in this thesis is designed to be adaptable, with the expectation that minor targeted retraining on locally collected data would be sufficient to achieve reliable performance across different urban contexts.

CSI Parameter Coverage. The current CSI framework incorporates three detected accessibility features, marked crossings, tram tracks, and traffic lights. The crosswalk length class derived from depth estimation is also considered. While these parameters capture a meaningful set of physical attributes relevant to crosswalk accessibility, they represent a subset of the full range of factors that influence crossing difficulty for mobility-impaired pedestrians. Previous studies have proposed broader evaluation frameworks incorporating additional parameters such as the number of vehicle lanes, the presence and width of dedicated bicycle lanes, the number of tram track lanes, pavement surface quality, and cross-slope gradient (Ahmed et al., 2021). Each of these attributes contributes independently to crosswalk severity and their inclusion would enhance the descriptive power of the index.

One notable example is the presence of dedicated bicycle lanes, which introduce an additional crossing hazard for mobility-impaired pedestrians. While the effect of bicycle lanes is partially reflected in the CSI through its contribution to overall crosswalk length, the specific risks associated with bicycle infrastructure are not independently represented as a severity factor. Cyclists traveling at speed in a dedicated lane present a meaningfully different risk profile compared to motorized vehicle lanes, particularly for mobility-impaired pedestrians. This distinction warrants explicit modeling in future extensions of the CSI. Similarly, the current treatment of tram track detections does not differentiate between single-directional tram service and more complex multi-lane configurations, such as bidirectional tracks or the multi-track layouts commonly found near tram stations and major intersections. A crosswalk overlapping a single tram track and one overlapping four tram lanes at a central interchange are treated equivalently in the current CSI. Incorporating lane-level tram track classification can substantially improve the CSI's ability to differentiate between crosswalk configurations of varying complexity and risk.

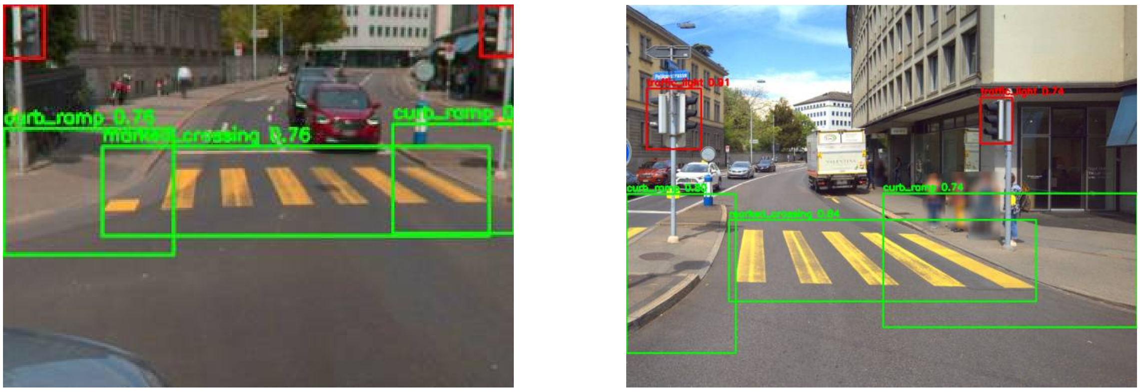
A core objective of this thesis is to develop a CSI computed entirely through an automated pipeline, without reliance on supplementary geospatial datasets or manual data collection. Incorporating additional parameters would therefore require the development and training of new detection tasks, which is feasible but beyond the scope of the thesis. The CSI framework presented in this thesis nonetheless demonstrates the viability of a fully automated approach to crosswalk severity assessment, providing a practical foundation upon which additional detection modules can be incrementally integrated to produce a more comprehensive accessibility indicator in future work.

Crosswalk Definition and Multi-Sectional Crosswalk. A further avenue for extending the current framework concerns the definition and delineation of crosswalk units, particularly in the context of long or multi-sectional crosswalks. In this thesis, a crosswalk is operationally defined as the pedestrian crossing segment between two directly connecting curb ramps. This mean that each section of a longer crossing is treated as an independent unit with its own CSI value. While this approach is appropriate for simple, single-span crossings, it does not fully capture the cumulative accessibility burden imposed by multi-section crosswalks. These long and multi-sectional crosswalk demand a pedestrian to traverse several sequential segments, each with its own physical characteristics and potential hazards, in order to complete a single crossing journey. Aggregating these individual sections into a composite crosswalk unit and computing a holistic CSI for the full crossing remains a meaningful direction for future development. It would better reflect the total navigational demand experienced by mobility-impaired pedestrians attempting to cross a wide or complex intersection.

However, this aggregation presents non-trivial detection and topological challenges that are difficult to resolve using street view imagery alone. There are mainly three types of multi-sectional crosswalks presence in the study area: crosswalk with sections separated by refuge island (Figure 28), crosswalk with sections separated by refuge island and lead to multiple directions (Figure 29), and crosswalk with sections separated by tram stop or tram station (Figure 30).

For the first type, where the multi-sectional crosswalk is simply separated by one or two refuge island, a possible solution would be train the model to detect refuge islands as an additional class. However, detecting the refuge island could be challenging for the current YOLO11s model, as most of the refuge islands are only partially visible within the SVI frames. From a CSI perspective, this type of multi-sectional crosswalk should arguably be assigned a lower severity than a continuous crosswalk of equivalent total length. Unlike an uninterrupted long crosswalk, the presence of a refuge island provides a stopping point that affords additional reaction time between crossing stages. This is a particularly important factor for mobility-impaired pedestrians with limited speed or stamina. This reasoning is intuitive but remains to be empirically validated, and further participatory survey data would be needed to confirm whether refuge islands are consistently perceived as reducing crossing difficulty in practice.

Instead of having the refuge island as a stopping point, it can be more functional in some multi-sectional crosswalks. In this type of crosswalk, the triangular refuge island divide the crosswalk into multiple directional sections that may lead to different destinations or require the pedestrian to re-orient between sections (see Figure 29). In such configurations, it is necessary to identify each possible crossing trajectory and assign a CSI to each distinct route, as the accessibility experience may differ substantially depending on the direction of travel and the specific sequence of sections encountered. Reliably inferring these topological relationships, including the spatial extent of refuge islands, the directionality of each crossing segment, and the connectivity between sections, from street view images alone is inherently challenging. The street view images capture only a partial and perspective-distorted view of the intersection geometry. A possible solution of these cases would be consider the spatial coordinates and open footpath data as a post-processing step, which is beyond this thesis.



(a) (b)

Figure 28. Example of sub-sectional crosswalks separated by a refuge island: (a) Section 1 (Crosswalk ID: c161_1) and (b) Section 2 (Crosswalk ID: c161_2).

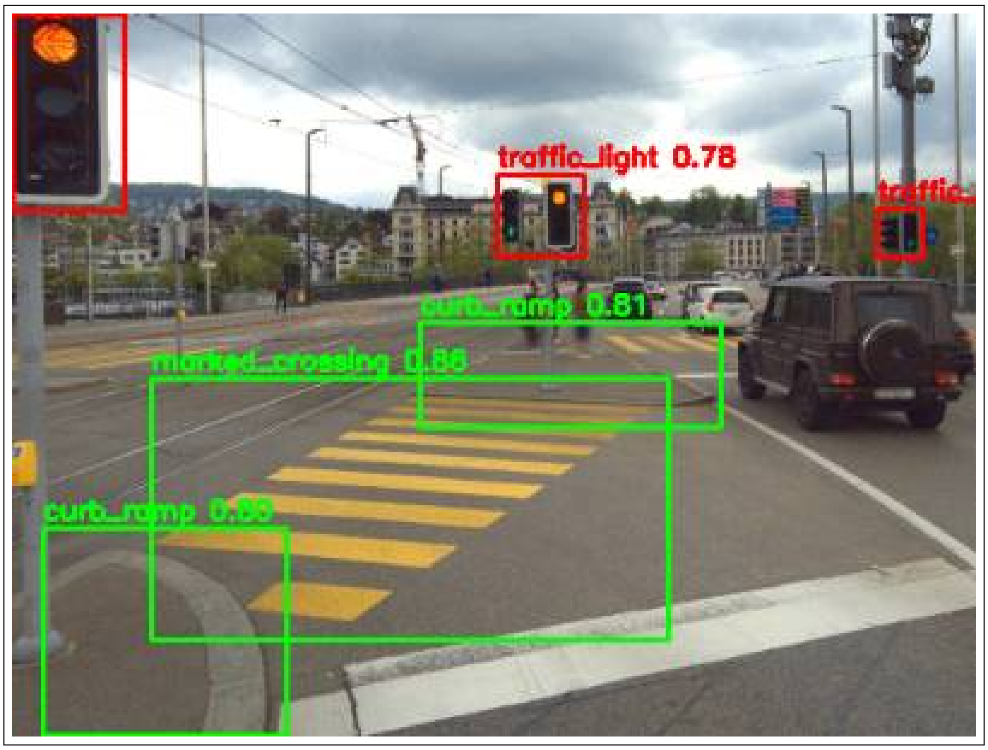
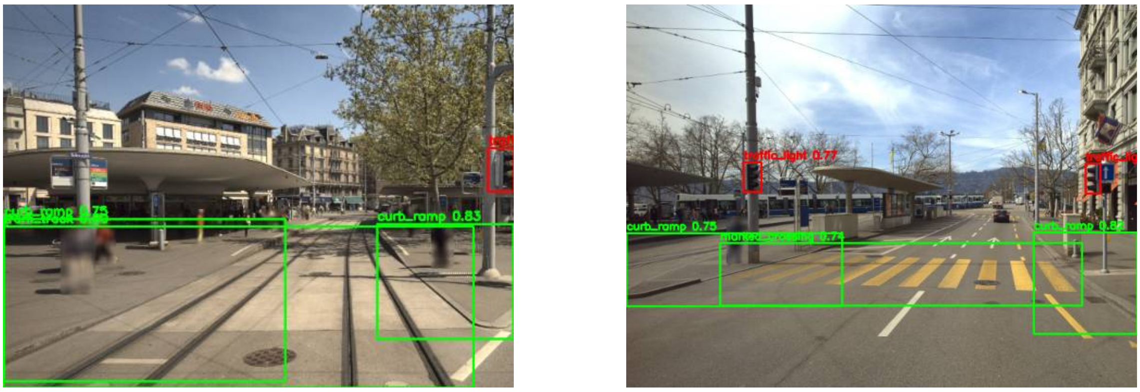


Figure 29. Example of crosswalks divided into multiple directional sections.

A third type of multi-sectional crosswalk involves sections separated by tram stops (see Figure 30). In such configurations, pedestrians have the option to exit the road at the tram stop or continue crossing to the far side of the road, resulting in multiple possible crossing trajectories within a single crosswalk. Ideally, each distinct crossing scenario should be treated as an independent crosswalk unit and assigned its own CSI value, as the accessibility experience differs substantially depending on the chosen route. As with the refuge island case, resolving these topological relationships solely from street view imagery is challenging, and the most viable solution would be the integration of additional geo-referenced data sources, such as tram stop locations and public transport network topologies, to reconstruct the full range of possible crossing trajectories.



(a) (b)

Figure 30. Example of long crosswalks separated tram stations: (a) Crosswalk ID: c305, and (b) Crosswalk ID: c306.

Overall, Addressing this limitation would likely require the integration of supplementary geospatial data sources, such as pedestrian footpath network data, public transport stop locations, and tram or bus route topologies. These data can be openly sourced from the city's data portal. Realizing this major limitation would represent a significant step towards a fully intersection-level accessibility assessment framework. Such a tool would be capable of capturing the complete crossing experience of mobility-impaired pedestrians across complex urban intersections, rather than evaluating individual segments in isolation. However, as stated in the objectives of this thesis, addressing this limitation using computer vision techniques applied solely to street view imagery remains a considerable challenge.

5.3. Further Outlooks

Automated Pipeline Reliability and the Role of Human Validation. While the proposed pipeline demonstrates the feasibility of a fully automated approach to crosswalk accessibility assessment, automated outputs alone are not sufficiently reliable to serve as the sole basis for infrastructure decision-making. Detection errors, depth estimation inaccuracies, and false positive classifications introduce uncertainties that accumulate through the pipeline and may affect the accuracy of individual CSI values. Cross-validation with human judgment is therefore necessary before the pipeline outputs are used to directly inform planning or investment decisions. Rather than replacing human expertise, the pipeline is best positioned as an initial screening or validation tool that rapidly identifies crosswalks of potential concern across a large area. After the utilization of the pipeline, targeted manual inspection or participatory survey can be applied to prioritize and verify the most critical cases. This hybrid approach of combining the scalability and objectivity of automated detecting with the contextual knowledge and lived experience of mobility-impaired pedestrians is consistent with best practices in inclusive urban planning. By integrating automated severity assessment as one component within a broader participatory planning framework, the pipeline can contribute to faster, more evidence-based, and more inclusive infrastructure improvement processes without bypassing the human dimension that is essential to meaningful accessibility evaluation.

Crosswalk Typology. Another potential extension of the current work is the integration of a crosswalk typology classification into the accessibility assessment pipeline. A preliminary typology framework is considered during the course of this thesis, with an initial set of categories proposed based on the physical and infrastructural characteristics of detected crosswalks (see Table 14). The typology approach is ultimately not pursued in the final implementation, as the classification criteria exhibited considerable conceptual overlap with the components already captured by the CSI, rendering it partially redundant within the current framework.

Table 14. Crosswalk Typology

ID	Crosswalk Typology
1	Marked crosswalk
2	Unmarked crosswalk
3	Marked crosswalk with tram track
4	Unmarked crosswalk with tram track

Nevertheless, the underlying rationale of crosswalk typology presents a meaningful avenue for future development. The core logic is that crosswalk types which occur less frequently within a given urban environment may pose greater accessibility challenges to pedestrians, particularly those with mobility impairments, precisely because of their rarity. Pedestrians who regularly navigate a city develop familiarity with the most common crosswalk configurations, and this repeating exposure builds a degree of implicit spatial and behavioral adaptation. Conversely, encountering an uncommon crosswalk type that deviates from the predominant urban pattern may introduce unexpected navigational demands, reduce environmental predictability, and heighten cognitive load. These challenges disproportionately affect mobility-impaired pedestrians, who may already be operating at the limits of their physical and perceptual capacity. Incorporating a frequency-weighted typology score into the CSI framework, where less common crosswalk types receive a higher severity contribution, could therefore enhance the index's sensitivity to this dimension of accessibility risk. In future works, this can provide a more nuanced characterization of crosswalk severity across diverse urban environments.

VR Participatory Survey. Beyond the methodological extensions discussed above, broader advancements in pedestrian behavior research offer promising directions for future accessibility assessment. Current approaches to collecting pedestrian accessibility data, ranging from participatory field surveys and online crowd-sourcing to automated machine learning pipelines, each carry inherent limitations in terms of scalability, data quality, or participant coverage. Virtual Reality (VR) presents an emerging alternative that is particularly well-suited to accessibility research involving mobility-impaired pedestrians (Feng et al., 2021). VR environments allow participants to experience and evaluate crosswalk configurations without the physical demands and safety risks associated with real-world field surveys, lowering the barrier to participation for wheelchair users, elderly pedestrians, and others with limited mobility (Feng et al., 2021). Compared to online crowdsourcing tools, VR-based data collection may require less raw image data as input. As the environment can be synthetically constructed and controlled, the usage of VR could still capture rich behavioral and perceptual responses from participants. Incorporating VR-based user evaluations as a complementary validation layer alongside the automated CSI pipeline could therefore enhance both the inclusiveness of the data collection process and the ecological validity of the resulting accessibility assessments.

6. Conclusion

This thesis addresses the challenge of assessing crosswalk accessibility for mobility-impaired pedestrians in urban environments through an automated, data-driven approach. Existing methods for crosswalk accessibility assessment have largely relied on participatory field surveys and crowd-sourcing platforms, which, while valuable, are limited in scalability, objectivity, and spatial coverage. To address these gaps, this thesis developed and assessed an end-to-end computer vision pipeline for detecting pedestrian accessibility features and computing a quantitative Crosswalk Severity Index (CSI) for District 1 in the City of Zurich.

The pipeline integrates three core components: a YOLO11s object detection model trained to identify curb ramps, marked crossings, tram tracks, and traffic lights from street view imagery; a ZoeDepth-based monocular depth estimation module for deriving crosswalk length estimates; and a multi-criteria CSI framework that combines detected features and estimated geometric measurements into a single severity score. The methodology is evaluated against ground truth annotations and validated through comparison with the ZuriACT crosswalk severity evaluation. The scalability of the approach is tested in a transferability analysis with an independent set of tram-mounted street view imagery.

The YOLO11s model achieves strong overall detection performance, with a mAP50 of 0.896 across all four classes. Marked crossings and tram tracks are detected with particularly high accuracy, benefiting from their visually distinctive and consistent appearance in the Zurich urban context. Curb ramps presents the greatest detection challenge, with a recall of 0.769 and a high background false negative rate, reflecting the visual variability and subtle appearance of this class. These results are broadly consistent with findings from previous studies, confirming that while deep learning-based detection is effective for visually salient accessibility features. Curb ramp detection remains an inherently difficult task that may benefits from larger and more targeted training datasets.

The CSI is computed for 476 crosswalks and normalized to a five-point scale for direct comparison with ZuriACT severity ratings. While the two systems exhibit different distributional profiles, both systems agree that severely inaccessible crosswalks constitute a small minority of the current pedestrian infrastructure. The CSI's more distributed profile is attributed to its integration of objective physical measurements alongside qualitative feature detection, capturing dimensions of accessibility that perception-based ratings alone may not fully reflect. The spatial distribution of CSI values reveals meaningful geographic patterns, with lower severity crosswalks clustered in residential areas and higher severity crosswalks concentrated along major tram corridors.

The transferability analysis applies the full pipeline to approximately 16,120 tram-mounted street view images from iNovitas 2024 data, successfully extracting 58 crosswalk instances. The severity profiles of the detected crosswalks are consistent with those of the main study, with normalized CSI values predominantly falling within intermediate severity levels. This supports the reproducibility and scalability of the pipeline within a similar urban context. Some challenges related to camera perspective, route directionality, and detection scale were identified as factors requiring further attention in future deployments.

All three research objectives are met. The thesis successfully demonstrates the feasibility of automating pedestrian accessibility feature detection using a lightweight deep learning model, developed a multi-criteria severity index that provides an objective and reproducible basis for crosswalk assessment, and validated the pipeline's transferability to an independent dataset. Together, these contributions advance the state of the art in automated urban accessibility assessment and provide a practical foundation for data-driven pedestrian infrastructure planning. Cross-validation with human judgment and participatory assessment remains an important step before the pipeline outputs are used directly to inform infrastructure planning decisions. The pipeline is therefore best understood not as a replacement for human expertise, but as a scalable and objective initial assessment tool that can rapidly identify priority locations for further investigation.

In conclusion, this thesis demonstrates that automated computer vision methods can provide a scalable, objective, and reproducible basis for crosswalk accessibility assessment. By moving beyond subjective perception-based ratings towards data-driven severity indicators grounded in physical infrastructure measurements, this work contributes empirical evidence towards the development of safer, more inclusive, and more equitable urban pedestrian environments for mobility-impaired pedestrians.

References

M. Aghaabbasi, M. Moeinaddini, M. Z. Shah, and Z. Asadi-Shekari. 2017. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *Journal of Transport & Health*, 5:97–112.

M. Aghaabbasi, M. Moeinaddini, M. Z. Shah, Z. Asadi-Shekari, and M. A. Kermani. 2018. Evaluating the capability of walkability audit tools for assessing sidewalks. *Sustainable Cities and Society*, 37:475–484.

T. Ahmed, M. Moeinaddini, M. Almoshaogeh, A. Jamal, I. Nawaz, and F. Alharbi. 2021. A new pedestrian crossing level of service (PCLOS) method for promoting safe pedestrian crossing in urban areas. *International Journal of Environmental Research and Public Health*, 18(16):8813.

H. Allahbakhshi. 2023. Towards an inclusive urban environment: A participatory approach for collecting spatial accessibility data in Zurich (short paper). In *12th International Conference on Geographic Information Science (GIScience 2023)*, pages 13–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

H. Allahbakhshi. 2025. Enhancing urban accessibility: Participatory Mapping and Technology for Inclusive Digital Maps. *Abstracts of the ICA*, 10:9.

H. Allahbakhshi and A. Ardüser. 2024. Navigation challenges in urban areas for persons with mobility restrictions (short paper). In *16th International Conference on Spatial Information Theory (COSIT 2024)*, pages 22:1–22:7. Schloss Dagstuhl–Leibniz-Zentrum für Informatik.

H. Allahbakhshi et al. 2026. ZuriACT: Zurich Accessible CiTy—Promoting Inclusive Urban Mobility in Zurich through Digitalization and Participatory Research. Submitted to the Journal of Urban Mobility.

Z. Asadi-Shekari, M. Moeinaddini, and M. Zaly Shah. 2013. Disabled pedestrian level of service method for evaluating and promoting inclusive walking facilities on urban streets. *Journal of Transportation Engineering*, 139(2):181–192.

Chiara Ballinari. 2025. Assessing Urban Accessibility Measurement Algorithms within the X-Minute City in Zurich. GEO 511 Master's Thesis, Department of Geography, University of Zurich, Zurich, Switzerland, January.

Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg.

S. Bennett, K. Kirchhoffer, W. C. Miller, and J. Garden. 2009. Wheelchair accessibility: Descriptive survey of curb ramps in an urban area. *Disability and Rehabilitation: Assistive Technology*, 4(1):17–23.

Shariq Farooq Bhat, Robert Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. 2023. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288*.

Yang Bian, Kun Liang, Xiaohua Zhao, Haijian Li, and Liping Yang. 2020. Evaluating the effectiveness of new-designed crosswalk markings at intersections in china considering vehicle-pedestrian interaction. *Accident Analysis & Prevention*, 139:105498.

Geoff Boeing. 2021. Off the grid... and back again? The recent evolution of American street network planning and design. *Journal of the American Planning Association*, 87(1):123–137.

S. N. Chowdhury, S. Dhawan, and A. Agnihotri. 2016. Crowd-sourcing for smart cities. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, pages 360–365, Bangalore, India. IEEE.

Department of Geography, University of Zurich. 2024. ZuReach: Zurich Urban Reachability & Accessibility Enhancement Through Digital Technology. Accessed: 2026-03-08.

Chyi-Ren Dow, Huu-Huy Ngo, Liang-Hsuan Lee, Po-Yu Lai, Kuan-Chieh Wang, and Van-Tung Bui. 2020. A crosswalk pedestrian recognition system by using deep learning and zebra-crossing recognition techniques. *Software: Practice and Experience*, 50(5):630–644.

M. Duan, S. Kiami, L. Milandin, J. Kuang, M. Saugstad, M. Hosseini, and J. E. Froehlich. 2022. Scaling crowd+ ai sidewalk accessibility assessments: Initial experiments examining label quality and cross-city training on performance. In

Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility, pages 1–5.

Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. 2012. Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200.

Y. Feng, D. Duives, W. Daamen, and S. Hoogendoorn. 2021. Data collection methods for studying pedestrian behaviour: A systematic review. *Building and Environment*, 187:107329.

Caroline Gallez and Benjamin Motte-Baumvol. 2017. Inclusive mobility or inclusive accessibility? A European perspective. *Cuadernos Europeos de Deusto*, (56):79–104.

A. I. Georgescu, H. Allahbakhshi, A. Darvishy, and R. Weibel. 2026. Same sidewalk, different experiences: How mobility restrictions shape perceptions of urban pedestrian infrastructure and accessibility. Unpublished manuscript.

GIS-Zentrum, Geomatik + Vermessung, Tiefbau- und Entsorgungsdepartement. 2026. Stadtkreise. Open Data Zürich, Accessed: 2026-03-28.

Robin Hambleton. 2015. *Leading the Inclusive City: Place-based Innovation for a Bounded World*. Policy Press, Bristol.

Julienne Hanson. 2004. The inclusive city: delivering a more accessible urban environment through inclusive design. *IIRA 2004 Tokyo Regional Conference*.

Jan Hauke and Tomasz Kossowski. 2011. Comparison of values of Pearson's and Spearman's correlation coefficients on the same sets of data. *Quaestiones Geographicae*, 30(2):87–93.

iNovitas AG. 2023. The new infra3D Release 3.1.5 is now live. Infra3d web-client, Accessed: 2026-03-08.

iNovitas AG. 2026a. infra3D service: Digital Street & Railways. Accessed: 2026-02-15.

iNovitas AG. 2026b. infra3D: The Smart Platform for Digital Roads and Railways. Accessed: 2026-02-15.

Dong Yeong Jeong, Jiyoung Kwahk, Sung H. Han, Joohwan Park, Mingyu Lee, and Hyeji Jang. 2018. A pedestrian experience framework to help identify impediments to walking by mobility-challenged pedestrians. *Journal of Transport & Health*, 10:334–349.

B. Raghuram Kadali and P. Vedagiri. 2015. Evaluation of pedestrian crosswalk level of service (los) in perspective of type of land-use. *Transportation Research Part A: Policy and Practice*, 73:113–124.

Murugeswari Kandavel, Aanjankumar Sureshkumar, and Rajesh Kumar Dhanaraj. 2025. Autonomous traffic-monitoring pedestrian crossing detection with motion sensors and maintenance decision systems in smart cities using yolov8. *Engineering Proceedings*, 118(92).

Efthimis Kapsalis, Nils Jaeger, and Jonathan Hale. 2024. Disabled-by-design: effects of inaccessible urban public spaces on users of mobility assistive devices – a systematic review. *Disability and Rehabilitation: Assistive Technology*, 19(3):604–622. Epub ahead of print 2022.

Rahima Khanam and Muhammad Hussain. 2024. Yolov11: An overview of the key architectural enhancements. *arXiv preprint arXiv:2410.17725*.

V. A. Krylov, E. Kenny, and R. Dahyot. 2018. Automatic discovery and geotagging of objects from street view imagery. *Remote Sensing*, 10(5):661.

Wenbo Lan, Jianwu Dang, Yangping Wang, and Song Wang. 2018. Pedestrian detection based on yolo network model. *Proceedings of 2018 IEEE International Conference on Mechatronics and Automation*, pages 1547–1551.

A. A. A. Lawal and T. Agbaje. 2026. Deep Learning on High-Resolution Satellite and Street-View Imagery for Automated Roadway Asset Classification and Pavement Condition Index (PCI) Prediction. *Preprint / Technical Report*.

C. Li, K. O. Y. Ma, M. Saugstad, K. Fujii, M. Delaney, Y. Eisenberg, N. Sikka, and J. E. Froehlich. 2024. “I never realized sidewalks were a big deal”: A Case Study of a Community-Driven Sidewalk Accessibility Assessment using Project Sidewalk. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–18, New York, NY, USA. ACM.

M. Li. 2021. Assisting Vision-Impaired Pedestrians To Cross Streets: An Innovative Acoustic Ranging Approach. *IEEE Transactions on Intelligent Transportation Systems*, 22(10):6334–6345.

X. Lu, Y. Xue, Z. Wang, H. Xu, and X. Wen. 2024. X-CDNet: A real-time crosswalk detector based on YOLOX. *Journal of Visual Communication and Image Representation*, 102:104206.

J. S. O'Meara, J. Hwang, Z. Wang, M. Saugstad, and J. E. Froehlich. 2025. RampNet: A two-stage pipeline for bootstrapping curb ramp detection in streetscape images from open government metadata. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6656–6665.

Y. R. Panta, S. Azam, B. Shanmugam, K. C. Yeo, M. Jonkman, F. De Boer, and M. Alazab. 2019. Improving accessibility for mobility impaired people in smart city using crowdsourcing. In *2019 Cybersecurity and Cyberforensics Conference (CCC)*, pages 47–55. IEEE.

Michael Prescott. 2020. *Towards a conceptual framework of navigation for people who use wheeled mobility devices in the community*. Ph.D. Dissertation, University of British Columbia, Vancouver, Canada.

Hannah Ritchie, Veronika Samborska, and Max Roser. 2024. Urbanization. Accessed: 2026-03-06.

M. Saha, M. Saugstad, H. T. Maddali, A. Zeng, R. Holland, S. Bower, et al. 2019. Project sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Norbert Schwarz. 1999. Self-reports: How the questions shape the answers. *American Psychologist*, 54(2):93–105.

Philip Sedgwick. 2014. Spearman's rank correlation coefficient. *BMJ*, 349:g7327.

A. P. Soares Müller, V. Goulart Dorneles, A. Ruiz-Padillo, and F. Vieira Romano. 2023. Sidewalk assessment from the perspective of accessibility: A systematic literature review. *Journal of Urban Planning and Development*, 149(3):04023032.

G. Spaias, V. Naserentin, N. Pitsianis, and A. Logg. 2025. Towards Large Scale Facade Feature Extraction Using Vision Transformers & Panoramic Street View Imagery. In *2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pages 1–8. IEEE.

Label Studio. 2015. LabelImg: A graphical image annotation tool and label object bounding boxes in images. Accessed: 2026-02-15.

C. Sundling and M. Jakobsson. 2023. How do urban walking environments impact pedestrians' experience and psychological health? A systematic review. *Sustainability*, 15(14):10817.

The Oxford Review. 2026. Mobility Impairment – Definition and Explanation. Accessed: 2026-03-07.

Tiefbauamt, Kanton Zürich. 2026. Strassennetz (OGD). Open Data Zürich, Accessed: 2026-03-28.

V. Tümen and B. Ergen. 2020. Intersections and crosswalk detection using deep learning and image processing techniques. *Physica A: Statistical Mechanics and its Applications*, 543:123510.

Ultralytics. 2024. Ultralytics yolo11.

United Nations. 2016. Good practices of accessible urban development. Technical report, Department of Economic and Social Affairs.

United Nations. 2024. Goal 11: Make cities and human settlements inclusive, safe, resilient and sustainable. Department of Economic and Social Affairs, Accessed: 2026-03-06.

Thorsten Völkel, Romina Kühn, and Gerhard Weber. 2008. Mobility impaired pedestrians are not cars: Requirements for the annotation of geographical data. In *International Conference on Computers Helping People with Special Needs (ICCHP 2008)*, volume 5105 of *Lecture Notes in Computer Science*, pages 1085–1092. Springer.

G. Weld, E. Jang, A. Li, A. Zeng, K. Heimerl, and J. E. Froehlich. 2019. Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 196–209.

Z. D. Zhang, M. L. Tan, Z. C. Lan, H. C. Liu, L. Pei, and W. X. Yu. 2022. CDNet: A real-time and robust crosswalk detection network on Jetson Nano based on YOLOv5. *Neural Computing and Applications*, 34(13):10719–10730.

Zürich Tourism. 2024. Zürich sans barrières. Accessed: 2026-02-14.

Zürich Tourism. 2024. The 12 districts of zurich. Accessed: 2024-05-22.

A. Personal Declaration

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

I used AI tools including ChatGPT, Gemini, and Claude for English proofreading.

Leqi Wang
29.03.2026 at Zurich