BirdCV-LiDAR: A Multimodal Data Fusion Framework for Automated Sidewalk Infrastructure Assessment

Chan Young Koh[1], Marwan Abdelatti[2], Abdeltawab Hendawi[1]

[1] University of Rhode Island, Kingston, Rhode Island 02881, USA - {ckoh04, hendawi}@uri.edu
[2] Providence College, Providence, Rhode Island 02918, USA - mabdelat@providence.edu

Keywords: Smart city, Spatial analysis, Sidewalk feature detection, Sidewalk feature analysis, LiDAR, Computer Vision

Abstract

Assessing sidewalk infrastructure for accessibility compliance is an important task in urban planning; however, traditional methods are often manual, subjective, and resource-intensive. This paper introduces BirdCV-LiDAR, a multimodal data fusion framework for an automated assessment of sidewalk infrastructure. The proposed approach integrates high-resolution bird’s-eye-view (HR-BEV) imagery with aerial LiDAR point cloud (ALPC) data to automatically detect, measure, and assess sidewalk features for compliance with accessibility standards. By combining YOLO-oriented bounding box (OBB) models with precise LiDAR-based elevation data, the framework enables accurate dimensional and slope evaluations of sidewalk features, such as crosswalks and truncated domes. Validation with a 12-inch (0.3048 m) inclinometer shows that LiDAR-based slope measurements achieve 84.7% accuracy, with a root-mean-square error (RMSE) of 0.1152 m for crosswalk width measurements. The framework achieves 81.0% accuracy in determining ADA-PROWAG compliance, providing an adaptable, expandable solution for improved urban accessibility assessments.

1. Introduction

In urban infrastructure assessment, the Pedestrian Level of Service (PLOS) is an important measure for promoting inclusivity and ensuring mobility rights for individuals with diverse physical abilities (Williams et al., 2005, Sadik-Khan, 2012). However, the correlation between a high PLOS value and the healthiness of pedestrian facilities remains questionable. Traditional approaches to measuring PLOS rely heavily on community surveys, a method with inherent subjectivity that may compromise pedestrian safety (Axelson et al., 1999, Huber et al., 2013).

The utilization of remote sensing technology has been pursued for decades, particularly for state and national land surveys (U.S. Geological Survey, 2025). Significant advancements emerged in remote sensing technologies, particularly in satellite and aerial imagery, alongside the progression of machine learning algorithms (Royimani et al., 2019). Recent advancements are creating new opportunities to simplify the assessment of sidewalk infrastructure, enabling the capture of large areas at once. Using these tools, researchers and urban planners can gain comprehensive insights into sidewalk conditions, identify areas that do not meet compliance standards, and prioritize remediation work.

Integrating multiple sensor modalities, known as multisensor data fusion (Hall and Llinas, 1997), enhances object detection by providing complementary perspectives and overcoming the limitations of individual sensors. Researchers widely recognize the potential of multisensor data fusion combining drone imagery and LiDAR data to enhance the precision and reliability of spatial analyses across diverse multidisciplinary applications (Jensen and Cowen, 1999, Schofield et al., 2022, Kyriou et al., 2023, Himeur et al., 2022). Given these advantages, aligning data from multiple sensors, particularly drone imagery and LiDAR, is challenging due to temporal-resolution discrepancies (Zhang et al., 2019). More recently, there has been a surge of works on object detection that fuse remote sensing datasets (Wang et al., 2025, Bai et al., 2022). The fusion of remote sensing datasets for object detection has shown great promise; however, it also faces limitations, particularly in detecting small objects. In general, publicly available, wide-coverage satellite imagery datasets are often unclear and noisy. Likewise, detecting small objects without supplemental data will be difficult (Hoyer, 2024, Ding et al., 2021). To address this limitation, some works have incorporated other datasets for data fusion (Wu et al., 2024, Liang et al., 2022).

Light Detection and Ranging (LiDAR) technology has emerged as a powerful tool for urban infrastructure assessment, offering precise three-dimensional spatial information that complements traditional imaging approaches. Mobile LiDAR systems mounted on vehicles have been successfully employed for automated detection and inventory of roadway assets, including traffic signs, road markings, and street furniture (Guan et al., 2014, Pu et al., 2011). These systems capture dense point clouds from street-level perspectives, enabling detailed geometric analysis of infrastructure elements. Terrestrial laser scanning (TLS) has demonstrated its effectiveness in producing accurate 3D models of building facades and urban structures (Armesto et al., 2009, Puente et al., 2013), although the labor-intensive nature of ground-based acquisition limits scalability for large-area assessments. Aerial LiDAR platforms offer advantages for wide-area coverage, and recent studies have explored applications in sidewalk infrastructure assessment (Balado et al., 2019, Balado et al., 2020). Notably, research on automated detection of podotactile paving from aerial LiDAR point clouds has shown promise for accessibility assessments (Balado et al., 2019), though challenges remain in detecting low-profile features with moderate-density public datasets. Comprehensive reviews of point cloud-based urban object detection (Arefi et al., 2020, Grilli et al., 2017) highlight ongoing challenges in feature extraction from aerial LiDAR, particularly for small-scale infrastructure elements. Our work addresses these limitations by fusing moderate-density publicly available ALPC data with high-resolution bird’s-eye-view imagery, combining the com-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-303-2026

plementary strengths of both modalities for accurate sidewalk feature detection and dimensional analysis.

Recent advances in computer vision have enabled highly accurate, automated dimension measurement across industries. Leveraging computer vision for this purpose can enable non-contact measurement of object dimensions, improving speed and reducing manual errors in real-world applications (Nogueira et al., 2023, Li, 2018). A study suggested using the ground sampling distance (GSD) of a drone or satellite (Li et al., 2021) for non-contact measurement of objects in two-dimensional (2D) aspect; however, the alignment task using GSD results in substantial inaccuracy since inexpensive commercial drones rely on barometric readings to estimate their altitude (Colomina and Molina, 2014). Some studies emphasize the importance of facility assessments, but they primarily focus on evaluating vehicular roadways.(Bhat et al., 2020, Fan et al., 2019, Hassan et al., 2022). Most existing methods struggle to accurately identify pedestrian facilities from large-scale remote sensing data because they require high-quality, fine-resolution datasets. While some studies have used street-level imagery to detect accessibility issues (Weld et al., 2019, Smith et al., 2013, Ivanchenko et al., 2008), these approaches often lack the top-down perspective needed for comprehensive dimensional analysis and are painstaking to scale across large urban areas. Other works have focused on the use of satellite imagery. The resolution is often insufficient to resolve small but important features, such as truncated domes, or to measure curb ramp slopes accurately. (Li et al., 2021).

Considering these limitations, this paper proposes a novel framework that utilizes high-resolution bird’s-eye-view (HR-BEV) imagery and aerial LiDAR point cloud (ALPC) data fusion to detect and measure the dimensions and slopes of pedestrian facilities. By combining HR-BEV drone imagery with high-quality ALPC data (approximately 18 million LiDAR data points), the framework achieves precise feature identification and accurate three-dimensional (3D) measurements of small sidewalk features such as crosswalks and truncated domes. Using ALPC data helps address altitude discrepancies often seen in consumer-grade drone barometric sensors. This correction yields more accurate ground sampling distance (GSD) calculations. The measurements are then compared to the standards outlined in the Americans with Disabilities Act (ADA) (Kent, 2023, US Department of Justice, 2015) and the Public Rights-of-Way Accessibility Guidelines (PROWAG) (Board, 2011) to assess compliance and identify areas for improvement.

The proposed approach is illustrated in Figure 2. The framework integrates three key inputs: HR-BEV images, SubRip Subtitle (SRT) files containing drone metadata, and ALPC data. The HR-BEV images are used to train an oriented bounding box (OBB) object detection model for feature identification and localization. The SRT file and ALPC data are fused to calibrate the drone’s altitude, enabling precise GSD calculations and accurate dimensional measurements. Detected features are georeferenced by converting pixel-based bounding boxes to geographic coordinates (latitude, longitude, elevation) and aligning them with the ALPC data. This multimodal fusion approach improves both detection and measurement precision.

Our approach enables remote evaluation of PLOS and identification of accessibility barriers, contributing to more inclusive and pedestrian-friendly urban environments. The framework makes several contributions. First, it introduces BirdCV-LiDAR, a multimodal fusion framework that addresses the limitations of traditional manual assessment methods by providing accurate, efficient, and scalable evaluation of sidewalk infrastructure. Second, it pioneers the application of fused computer vision and LiDAR technologies for ADA-PROWAG compliance assessments of features such as truncated dome tiles and crosswalk slopes. Third, the framework is model-agnostic, enabling flexible selection of object detection models for diverse applications. Fourth, rigorous experiments evaluate multiple YOLO-OBB model variants, demonstrating effectiveness in detecting and measuring sidewalk features. Finally, extensive ground-truth validation with physical measurements verifies the framework’s applicability for real-world ADA-PROWAG compliance assessment. The framework addresses numerous ADA-PROWAG requirements, including slope and dimensional compliance, and provides an adaptable foundation that can be customized with additional sensors and analytical modules.

2. Materials and Methods

2.1 Study Area

The study was conducted on a large public university campus with a diverse range of pedestrian infrastructure. We surveyed approximately 890 acres of the campus, where a variety of sidewalk types, crosswalks, curb ramps, and other features were representative of a typical urban environment. Figure 1 illustrates the declared experimental boundary, where an aerial drone was deployed throughout the red polygonal area to capture high-resolution imagery. The corresponding ALPC data for this boundary are shown in greyscale, demonstrating the comprehensive elevation coverage available for the study area.

(a) Aerial drone deployment boundary

(b) ALPC data in greyscale

Figure 1. The declared boundary to experiment and evaluate the BirdCV-LiDAR framework

2.2 Data Acquisition

Our framework uses two primary data sources: high-resolution bird’s-eye-view (HR-BEV) imagery and aerial LiDAR point clouds (ALPC). HR-BEV imagery was captured using a DJI Mini 4 Pro drone at 4K resolution in a nadir (90-degree downward) view orientation. ALPC data was sourced from the USGS 3D Elevation Program (3DEP), comprising 61.1 billion points with an average nominal pulse spacing (ANPS) of 0.35 m, a density of 8 points/m^2, and a non-vegetated area (NVA) RMSEz of 2.17 cm. Each drone flight also produces a SubRip Subtitle (SRT) metadata file that encodes GPS coordinates, relative altitude, absolute altitude, and timestamps for every recorded frame, which serves as the backbone for spatial alignment between the 2D imagery and the 3D ALPC data. Our study uses $\approx$ 18 million ALPC data points to assess the declared boundary.

2.2.1 High-Resolution Aerial Drone Imagery In this study, we use a drone-agnostic method to record 4K videos,

capturing high-resolution footage of sidewalk segments and other components from a 90-degree downward (nadir) view. This is crucial for object detection and feature interpretation, which are essential for ADA-PROWAG compliance assessment.

2.2.2 Conversion to Images Post-flight, the captured video is processed into individual frames, each representing a still image in 4K resolution. High-resolution imagery is beneficial because many sidewalk components are small and can only be observed with detailed images. Each image frame retains the original video’s embedded GPS metadata, which encodes the geolocation of the frame’s center, facilitating precise spatial alignment with other data in a later stage.

2.2.3 Total Dataset Size and Coverage We process 100 HR-BEV sidewalk images, yielding 300 annotated assessment features (175 truncated dome tiles and 125 crosswalks), to create a balanced dataset for robust model training, while accounting for practical annotation and processing time constraints. Sidewalk partial segments were included as a third training class to reduce false positives, but are not the primary assessment targets of the framework. Each image is paired with an SRT file that contains GPS information, timestamps, relative altitude, absolute altitude, and drone-camera information, thereby adding spatial and temporal context to each image. This information serves as the backbone for data fusion, linking 2D data with 3D data, such as ALPC data.

2.3 Data Preprocessing

2.3.1 Image Patching We divided each 4K image (3840 x 2160 pixels) into four non-overlapping patches of 1920 x 1080 pixels to overcome hardware memory limitations during model training. This spatial tiling approach maintains the original aspect ratio in each patch, ensuring that objects within individual patches remain undistorted. Importantly, each patch preserves the original pixel-to-ground correspondence (GSD), which is necessary for accurate data fusion between 2D imagery and 3D ALPC data.

2.3.2 Data Annotation and Split Images are annotated to label two major sidewalk features at intersections: crosswalks and truncated domes, which connect curb cuts and crosswalks. There are existing annotation methods; however, we manually annotate the images since their dimensions are used for ADA-PROWAG compliance analysis. This process ensures that the model receives precise, human-verified feature data, which is important because sidewalk features can often be obscured or vary in appearance. We performed a stratified dataset split to ensure a fair distribution of features before training the models.

2.4 Computational Requirements and System Specifications

The BirdCV-LiDAR framework was implemented and evaluated on a workstation equipped with dual NVIDIA RTX 4090 GPUs (24GB VRAM each), an AMD Ryzen Threadripper PRO 5975WX processor (32 cores, 64 threads), and 128GB DDR4 RAM, running Ubuntu 22.04 LTS. The deep learning framework utilized PyTorch 2.0 with CUDA 11.8 for GPU acceleration.

Training Phase: YOLOv8m-obb model training required approximately 12 hours for 300 epochs on a dataset of approximately 100 annotated 4K images (divided into 400 patches of 1920×1080 pixels), with a batch size of 16. Peak GPU memory usage during training was approximately 18GB per GPU. The model was trained using data augmentation techniques, including random rotation, scaling, and color jittering, to improve robustness.

Inference Phase: For a typical 4K video frame (3840×2160 pixels) divided into four 1920×1080 patches, the inference time per frame was approximately 0.15 seconds, enabling near-real-time processing at 6-7 frames per second. The ALPC data fusion and slope calculation steps added approximately 0.3 seconds per detected feature, depending on the density of LiDAR points in the region of interest.

Storage Requirements: The framework requires approximately 50GB of storage for the trained model weights, preprocessed imagery, ALPC data, and intermediate results for the study area (890 acres). Raw 4K video footage requires approximately 1GB per minute of flight time.

Software Dependencies: The framework relies on open-source libraries including OpenCV 4.8 for image processing, GDAL 3.6 for geospatial data handling, scikit-learn 1.3 for kriging interpolation, and Ultralytics YOLOv8 for object detection. All dependencies are freely available and can be installed via standard package managers.

Scalability Considerations: The computational requirements scale linearly with the study area size. For a typical city block (approximately 10 acres), the processing time from raw video to final compliance assessment is approximately 30 minutes. The framework can be parallelized across multiple GPUs to process larger areas more quickly. For municipalities without access to high-end GPUs, cloud computing services (e.g., AWS, Google Cloud) can be used on a pay-per-use basis, with estimated costs of $5-10 per square kilometer of assessed area.

2.5 BirdCV-LiDAR: Multimodal Framework

Our BirdCV-LiDAR framework integrates HR-BEV imagery and ALPC data to enable the automated assessment of sidewalk infrastructure. The workflow of our framework is illustrated in Figure 2.

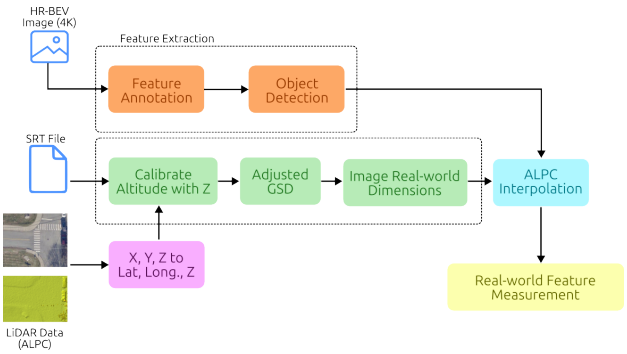

Figure 2. BirdCV-LiDAR framework workflow

2.5.1 Sidewalk Feature Detection: YOLO-OBB We train three YOLO model variants, namely YOLOv8m-obb, YOLO11m-obb, and YOLO12m-obb. These models are chosen for their single-stage, oriented bounding box (OBB) detection capabilities, which are important for our application. Unlike standard horizontal bounding boxes, OBB models provide rotated bounding boxes that tightly fit the actual orientation of

sidewalk features such as crosswalks and truncated dome tiles. This precise geometric representation is essential for accurate dimensional analysis, as it enables direct extraction of feature length, width, and orientation without the geometric ambiguity introduced by axis-aligned boxes when fused with the ALPC dataset. The models represent a progression in architectural complexity. YOLOv8m-obb introduces a decoupled head that separates classification and regression, improving performance in dense object detection. YOLO11m-obb and YOLO12m-obb further refine the architecture by incorporating attention mechanisms and more efficient backbones, thereby theoretically enhancing the ability to handle occluded, partially visible, or small-scale features.

2.5.2 Tackling Uncertain Drone Altitude A fundamental challenge in drone-based measurements is altitude uncertainty, as most lightweight drones rely on relative rather than sensor-based absolute altitude measurements. An illustration of the problem is depicted in Figure 3. The relative altitude is defined with the drone's takeoff location as zero altitude.

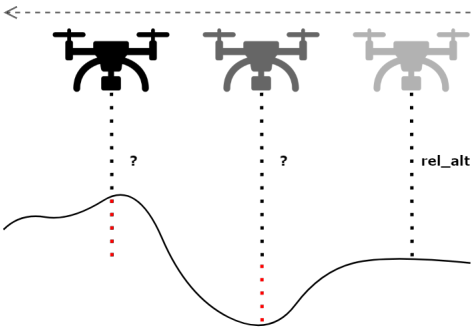

Figure 3. Incorrect altitude metric on a lightweight drone

The drone's relative altitude h_d is calibrated using LiDAR ground elevation h_L to find the corrected altitude h_c

$$h_c = h_d + (h_{L,drone} - h_{L,start}) \tag{1}$$

Then, the GSD is computed as:

$$GSD = \frac{Sensor}{f} \times h_c \tag{2}$$

where Sensor is the sensor pixel size and f is the focal length of the drone's camera.

Altitude calibration is necessary. By comparing the drone's initial altitude reading at its takeoff position with the precise elevation data from the ALPC at the same geographic coordinate, a baseline offset is established. This offset is then applied to all subsequent altitude readings from the drone's SRT file. This process effectively corrects for barometric drift and yields a more accurate altitude profile over the entire flight path, which is essential for precise GSD calculation. With calibrated altitude, GSD is reliably calculated for each image. For example, at a specific geographic location, if the relative altitude of a drone is 150 m (492 ft) and the elevation value of ALPC is 50 m (164.042 ft), we use Equations 1 and 2 to yield a GSD of 0.58398 cm/pixel, allowing pixel-based measurements to reflect real-world dimensions accurately.

2.5.3 Remote ADA-PROWAG Assessment

Pixel to Geocoordinate Conversion and Feature Dimension Validation. For each detected bounding box B_i, we compute its four corner points in real-world coordinates:

$$(X_i, Y_i) = transform(x_i, y_i, GSD) \tag{3}$$

Once converted, we determine the bounding box dimensions by computing the lengths of its sides. The longer side is designated the length (L), and the shorter side is designated the width (W). To ensure compliance with ADA-PROWAG standards, we validate the dimensions for different sidewalk features. The compliance condition for truncated domes is W $\geq$ 0.6096 m (2 ft) and L $\leq$ 2.0% in terms of change of slope, and for crosswalks is W $\geq$ 1.8288 m (6 ft). The feature is flagged as non-compliant if these conditions are not met.

Slope Assessment. This final step of our framework assesses the slope trend within and between extracted bounding boxes. The bounding box serves as the reference geometry for slope calculation. The process involves analyzing elevation changes across the boundary box between the features. Slope assessment is an important component of evaluating sidewalk accessibility compliance. This study focuses on sidewalk guidelines in PROWAG and examines the slopes and running slopes of detected features (Figures 4 and 5).

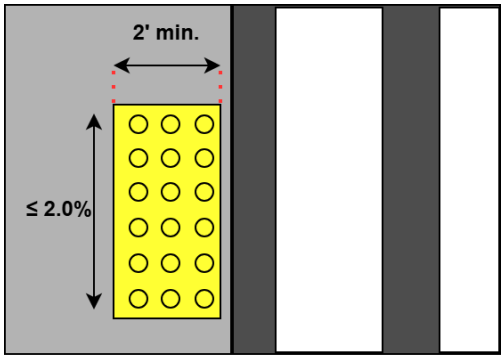

Figure 4. PROWAG R305.1.4 - Detectable surface size

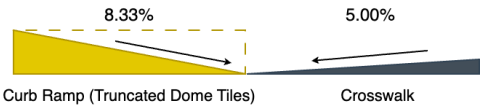

Figure 5. Running slope guidelines in PROWAG Ch. R304

During slope assessment, two slope calculations are involved: slope within one sidewalk feature (intra-slope) and slope between two sidewalk features (inter-slope). We begin this process by extracting the LiDAR points P_B for each bounding box region, such that:

$$P_B = \{p \in P | (X_{min}, Y_{min}) \leq (X_p, Y_p) \leq (X_{max}, Y_{max})\} \tag{4}$$

We evaluate compliance by checking the slopes of the features. Specifically, we analyze the slope behavior of each feature by computing elevation changes within its bounding box. An inter-slope analysis is performed between two distinct features sep-

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-303-2026 | © Author(s) 2026. CC BY 4.0 License.

arated by 0.9144 m (3 ft) to assess connectivity and ADA-PROWAG compliance.

2.6 Ground-Truth Validation Protocol

To ensure the accuracy of our computational slope analysis, we developed a systematic validation protocol using a 12-inch (0.3048 m) digital inclinometer. The ground-truth validation protocol involved systematic physical measurements using a 12-inch (0.3048 m) digital inclinometer (Smart Tool Levels ADA Slope Walker) with $\pm 0.1^\circ$ accuracy. For each detected feature, the validation team followed a standardized procedure: (1) Navigate to the feature location using GPS coordinates from the framework output; (2) Verify feature identity (crosswalk or truncated dome) and record any discrepancies; (3) Measure running slope along the centerline at 0.3048 m (12-inch) intervals, recording each measurement; (4) Measure cross slope perpendicular to the running slope direction; (5) For crosswalks, measure width and length using a measuring tape; (6) For truncated domes, count the number of domes and measure tile dimensions; (7) Photograph each feature with measurement equipment visible for documentation. This process required approximately 5-8 minutes per feature, and the validation of 247 features was completed over three weeks. Weather conditions during validation were recorded to ensure consistency with the conditions under which drone imagery was acquired. This validation protocol enables direct comparison between computational and physical slope measurements, providing confidence intervals for our ADA-PROWAG compliance assessments.

3. Results

3.1 Detection Model Performance

We evaluated the performance of three YOLO OBB models for sidewalk feature detection: YOLOv8m-obb, YOLO11m-obb, and YOLO12m-obb. The performance metrics for each model are presented in Table 1.

Table 1. Detection Model Performance Metrics

Model	Precision	Recall	mAP50	mAP50-95
YOLOv8m-obb	0.792	0.833	0.886	0.727
YOLO11m-obb	0.830	0.816	0.754	0.555
YOLO12m-obb	0.898	0.667	0.766	0.447

YOLOv8m-obb achieved the best overall performance, with a precision of 0.792, a recall of 0.833, and a mAP50 of 0.886. Based on these results, we selected YOLOv8m-obb as the optimal model for our framework. Figure 6 shows a visual comparison of the detection results from the three models, and Figure 7 shows more results using YOLOv8m-obb in various locations within the point of interest.

3.2 Dimension Measurement Accuracy

We assessed the accuracy of our framework’s dimension measurements using the root-mean-square error (RMSE) metric. The RMSE values for each YOLO model are shown in Table 2. Our empirical evaluation revealed that YOLOv8m-obb achieved superior performance compared to its successors, suggesting that architectural complexity does not always translate to better performance for domain-specific applications. This finding denotes the importance of empirical model selection rather than relying solely on architectural advancements.

(a) YOLOv8m-obb

(b) YOLO11m-obb

(c) YOLO12m-obb

Figure 6. YOLO OBB model comparison

Figure 7. YOLOv8m-obb results in different locations.

3.3 Slope Assessment Results

Our computational slope analysis was validated against inclinometer measurements. The intra-feature slope validation showed an accuracy of 84.7% with an error of ±2.0%. The inter-feature slope validation had a mean absolute error of 1.87%. Figure 8 shows the slope validation process using a physical 12-inch (0.3048 m) inclinometer.

3.4 Compliance Determination

We evaluated our framework’s ability to classify sidewalk features as ADA-PROWAG compliant or non-compliant by manually inspecting 100 detected features (50 crosswalks and 50 truncated domes). Given our state university campus setting, which has high ADA compliance standards, most features were compliant. Our framework achieved an overall accuracy of 81.0% in determining ADA-PROWAG compliance, with 74 true positives for compliant features, 7 true positives for non-compliant features, 12 false positives, and 7 false negatives.

Figure 8. Slope validation process using a physical 12-inch (0.3048 m) inclinometer. Using the inclinometer, we inspect intra-feature slope (left) and inter-feature slope (right) for selected sidewalk features in our point of interest.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-303-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 2. RMSE of Dimension Measurement (m)

Model	TD Width	TD Length	CW Width	CW Length
YOLOv8m-obb	0.1374	0.1202	0.1152	0.1214
YOLO11m-obb	0.2976	0.2782	0.3302	0.2991
YOLO12m-obb	0.2697	0.2601	0.2367	0.2354

The precision for non-compliant features was 36.8%, and the recall was 50.0%. The negative predictive value for compliant features (86.0%) demonstrates reliable identification of compliant infrastructure. False negatives resulted from subtle slope violations near compliance thresholds, while LiDAR noise or YOLO OBB model detection inaccuracies caused false positives. Figure 9 provides a map visualization of the compliance determination results, highlighting compliant and non-compliant features.

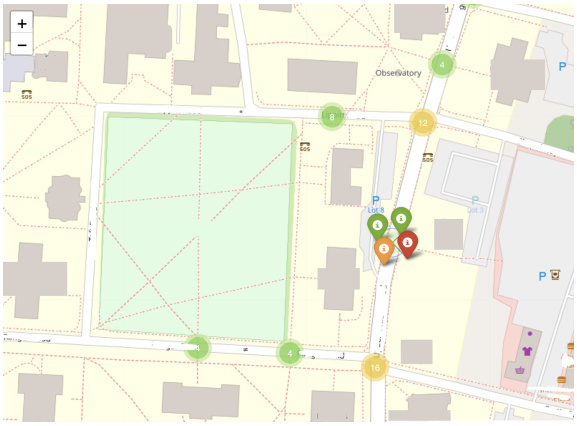

Figure 9. Compliance Visualization Map via OpenStreeMap (OpenStreetMap contributors, 2017)

3.5 Quantitative Comparison

To evaluate the effectiveness of our multimodal data fusion approach, we conducted a quantitative comparison of our complete BirdCV-LiDAR framework against approaches using only ALPC data or only HR-BEV images. Table 3 presents a quantitative comparison of the three approaches across key performance metrics.

The ablation study demonstrates the superior performance of our multimodal framework compared to single-modality approaches.

4. Discussion

4.1 Interpretation of Results

Our results demonstrate the effectiveness of our BirdCV-LiDAR framework for automated assessment of sidewalk infrastructure. The high accuracy of our slope measurements (84.7%) and the low error in our dimensional measurements (RMSE of 0.1152m for crosswalk width) highlight the benefits of our multimodal data fusion approach. Selecting YOLOv8m-obb as the detection model was crucial for achieving these results, as it provided the best balance of precision and recall for detecting small, often-obscured sidewalk features. The ablation study further underscores the value of our multimodal approach. The HR-BEV-only approach, although fast, cannot perform slope analysis and is prone to high-dimensional errors due to the lack of precise altitude data. The ALPC-only approach, although capable of accurate slope analysis, is prohibitively slow and labor-intensive because it requires manual feature detection. Our BirdCV-LiDAR framework successfully combines the strengths of both modalities, achieving high accuracy in dimensional and slope analysis while maintaining high automation and efficiency. The integration of these components is pivotal to the successful implementation of the framework.

4.2 Comparison with Existing Approaches

Compared to traditional manual methods for sidewalk assessment, our framework offers greater efficiency, objectivity, and scalability. The processing time of 6.3 minutes per acre is a substantial improvement over the estimated 300 minutes per acre for a manual ALPC-only approach. Furthermore, our automated approach eliminates the subjectivity inherent in manual assessments, leading to more consistent and reliable results. Compared with other automated approaches, our framework stands out. Many existing methods focus solely on either image-based detection or LiDAR-based analysis. Our framework, by contrast, integrates both, enabling a more holistic assessment. For example, while some studies have used deep learning on street-level images to identify accessibility issues, they often cannot perform accurate dimensional or slope analysis. Our top-down perspective, combined with LiDAR’s precision, enables a more comprehensive evaluation of ADA-PROWAG compliance.

Table 4 provides a detailed comparison of our BirdCV-LiDAR framework with existing sidewalk assessment approaches across multiple dimensions. As shown, our multimodal fusion approach addresses the key limitations of existing methods. While manual inspection remains the gold standard for regulatory compliance, it is not scalable for city-wide assessments. Street-level imagery approaches, such as those using Google Street View, suffer from oblique viewing angles and occlusions that make accurate dimensional measurements difficult. Satellite imagery, even at high resolution, cannot reliably detect small features like truncated domes or measure curb ramp slopes with sufficient precision.

LiDAR-based approaches offer high accuracy but are subject to practical limitations. Mobile LiDAR systems provide excellent detail but require expensive acquisition campaigns and are limited to road-accessible areas. Terrestrial laser scanning offers the highest accuracy but is extremely labor-intensive and not practical for large-scale deployment. Single-modality approaches using either ALPC or HR-BEV data alone are insufficient. ALPC-only methods struggle to detect small, low-profile features due to moderate point density (8 points/m^2), while HR-BEV-only methods cannot accurately measure slopes without reliable elevation data. Our fusion approach combines the complementary strengths of both modalities: HR-BEV imagery provides high-resolution feature detection, while ALPC data enables accurate 3D measurements. Critically, both data sources are publicly available (USGS 3DEP for ALPC, consumer drones for HR-BEV), making our approach scalable and cost-effective for municipal deployment.

4.3 Practical Implications

The BirdCV-LiDAR framework has several practical implications for urban planning and infrastructure management. It can be used to conduct large-scale assessments of sidewalk accessibility, identify noncompliant areas with ADA-PROWAG standards, and prioritize areas for improvement. The framework can

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-4-2026-303-2026 | © Author(s) 2026. CC BY 4.0 License.

Table 3. Comparison of BirdCV-LiDAR with single-modality approaches

Metric	BirdCV-LiDAR (Ours)	ALPC-Only	HR-BEV-Only
Feature Detection	Automated	Manual	Automated
Detection Precision	0.792	0.43	0.792
Detection Recall	0.833	0.38	0.833*
Slope Analysis	Yes	Yes	No
Slope MAE (inter-feature)	1.87%	2.34%	N/A
Dimension Accuracy	High	Low	Medium
Dimension RMSE (CW Width)	0.1152m	N/A	0.2367m
Processing Time/Acre	6.3 min	300 min	4.8 min
Scalability	High	Low	Medium
Labor Requirement	Low	Very High	Low

*HR-BEV-Only detection metrics are identical to BirdCV-LiDAR because both use the same YOLOv8m-obb model; the distinction lies in measurement accuracy, not detection performance.

Table 4. Comprehensive comparison of sidewalk assessment approaches

Approach	Data Source	Key Characteristics	Scalability
Manual Inspection	Inclinometer, tape measure	(+) High accuracy, regulatory accepted; (–) Labor intensive, time consuming, not scalable	Very Low
Street Imagery	Google Street View, mobile cameras	(+) Wide coverage, existing datasets; (–) Oblique angles, occlusions, no elevation data	Medium
Mobile LiDAR	Vehicle-mounted scanners	(+) Very high density, accurate 3D; (–) Expensive, road-accessible areas only, not public	Low
ALPC-Only	Public aerial LiDAR	(+) Free, wide coverage, accurate elevation; (–) Cannot detect small features, no texture	High
HR-BEV-Only	Drone imagery	(+) High resolution, flexible acquisition; (–) Altitude errors, no elevation, 2D only	Medium
BirdCV-LiDAR (Ours)	**HR-BEV + ALPC fusion**	**(+) Combines strengths, uses public data, accurate 3D, cost-effective; (–) Requires fusion, limited by ALPC density**	**High**

also be used to monitor changes in sidewalk infrastructure over time, enabling a more proactive approach to maintenance and repair. By providing a more accurate and comprehensive understanding of sidewalk accessibility, our framework can help cities create more inclusive and equitable urban environments.

4.4 Scalability and Generalizability

The framework's reliance on publicly available data sources is a key factor in its scalability and generalizability. The ALPC data used in this study is from the USGS 3D Elevation Program (3DEP), which provides high-quality LiDAR data across the entire United States (U.S. Geological Survey, 2025). Similarly, Natural Resources Canada provides extensive QL1 and QL2 aerial LiDAR data through its CanElevation series, which covers major Canadian cities (Natural Resources Canada, 2025). The availability of these datasets suggests that the BirdCV-LiDAR framework can be readily deployed in numerous urban areas across North America without the need for expensive, proprietary data collection. This makes it a highly adaptable solution for municipalities and government agencies seeking to conduct large-scale accessibility assessments.

4.5 Sensitivity Analysis and Robustness

LiDAR Point Density: The framework's performance is influenced by ALPC point density. Our study area utilized USGS 3DEP QL2 data with approximately 8 points/m^2. Based on our observations, slope accuracy is expected to degrade when point density falls below 4 points/m^2, as insufficient ground returns reduce the reliability of elevation interpolation within bounding-box regions. Conversely, QL1-standard densities above 12 points/m^2 are expected to provide only marginal accuracy improvements for slope estimation at the feature scale. These observations suggest that QL2 data represent a practical balance between data availability and accuracy for most municipal deployment scenarios.

Flight Altitude and GSD: Corrected drone altitude directly affects GSD and detection accuracy. In our study, flights were conducted at a relative altitude of approximately 30-50 m above the takeoff point. Because this relative altitude is referenced to the launch location, the effective height above ground varied across the study area and was corrected using ALPC elevation before GSD estimation. Accordingly, GSD was computed on a per-image basis rather than being assumed constant throughout our experiment. Within this operating range, flights near the lower end provide finer spatial resolution that can improve the visibility of small sidewalk features, such as truncated domes, but at the cost of reduced area coverage, longer mission time, and increased data volume. Flights near the upper end of the range improve coverage efficiency, but the coarser spatial resolution may slightly reduce detection precision for small features. The selected 30-50 m relative flight range, therefore, represented a practical trade-off among image resolution, area coverage efficiency, and processing burden.

Bounding Box IoU Threshold: In YOLO-based OBB detection, the Intersection over Union (IoU) threshold governs how tightly a predicted bounding box must overlap the annotated ground-truth box to be counted as a true positive during evaluation. Given that accurate dimensional measurements depend directly on the precision of the predicted bounding box geometry, we applied a strict IoU threshold of 0.80 during model

evaluation. This requires at least 80% overlap between predicted and ground-truth oriented bounding boxes, ensuring that only detections with near-exact geometric alignment are accepted. This conservative criterion was deliberately chosen to reflect the measurement-critical nature of the task: a loosely fitting bounding box would introduce systematic error into the pixel-to-real-world dimensional conversion, directly affecting compliance assessments.

Environmental Conditions: Weather conditions significantly affect both drone imagery quality and safe flight operations. Contrary to the assumption that pristine sunny days are ideal, our field experience found that overcast days with low wind speeds provided the most consistent imagery, as uniform cloud cover eliminates harsh shadows from buildings and trees that can obscure sidewalk features and increase false negatives. Rainy or snowy conditions are unsuitable for lightweight consumer drones due to both hardware limitations and degraded image quality. Wind speed is a critical operational constraint, as lightweight drones such as the DJI Mini 4 Pro are susceptible to instability in moderate to high winds, compromising nadir image alignment. Seasonally, data acquisition between late winter and early spring — before vegetation fully emerges — is strongly recommended. Dense foliage from trees and shrubs can occlude sidewalk features in imagery, directly reducing detection model performance. Leaf-off conditions also improve the accuracy of ALPC ground-point classification in vegetated areas, yielding more reliable elevation data for slope analysis.

4.6 Deployment Guidance and Operational Considerations

Data Availability Assessment: Before deployment, municipalities should verify ALPC data availability and quality through USGS Earth Explorer or Natural Resources Canada’s Open Maps portal. Key parameters to check include: (1) Point density (minimum 4 points/m^2 recommended); (2) Data vintage (ideally within 3-5 years); (3) Classification quality (ground points properly classified); (4) Vertical accuracy (RMSE $<$ 10cm for QL2). Data vintage is particularly important because urban environments change substantially over time. Figure 10 illustrates this with aerial imagery of the same campus location captured in 2012 and 2023, showing significant changes in infrastructure, vegetation, and surrounding development. Using outdated ALPC data in such areas risks misclassifying or missing features that have been added, removed, or modified since acquisition.

(a) 2012 (b) 2023

Figure 10. Temporal comparison of the same campus location in 2012 and 2023, illustrating how infrastructure and surroundings can change substantially over time, underscoring the importance of verifying ALPC data vintage before deployment.

Operational Workflow: Recommended deployment workflow consists of four phases: (1) Planning phase: Identify study area boundaries, verify ALPC coverage, obtain necessary drone flight permits; (2) Data acquisition phase: Conduct drone flights at 30-50 m relative altitude with 70% forward and 30% side overlap, record 4K video at 30fps; (3) Processing phase: Convert video to frames, run YOLOv8m-obb detection, perform data fusion and compliance analysis (estimated 6-8 hours processing time per square kilometer on recommended hardware); (4) Validation phase: Ground-truth a statistically significant sample (minimum 30 features) using physical measurements.

Model Generalization: The model was trained exclusively on a university campus environment, and its performance in other urban contexts has not been formally evaluated. Domain shift is a known challenge in deep learning-based detection, and differences in pavement materials, marking styles, and feature geometry across deployment environments are expected to affect detection accuracy. We recommend fine-tuning the YOLOv8m-obb model with locally annotated images from the target deployment area to adapt to site-specific infrastructure characteristics before operational use.

Regulatory Integration: For official ADA-PROWAG compliance reporting, we recommend using the framework as an initial screening and prioritization tool, followed by targeted physical validation of features flagged as non-compliant. This hybrid approach concentrates field inspection effort on the subset of features most likely to require remediation, rather than requiring exhaustive manual surveys of entire study areas. The framework’s outputs can be integrated into existing asset management systems via standard GIS formats (Shapefile, GeoJSON, KML), thereby facilitating their incorporation into municipal infrastructure workflows.

Cost-Benefit Analysis: The BirdCV-LiDAR framework was developed and validated in an academic research context, and a precise commercial cost estimate for a full deployment cycle is beyond the scope of this study. However, indicative cost components can be outlined. The drone hardware used in this study (DJI Mini 4 Pro) retailed at approximately $1,200 at the time of purchase, representing a low-cost entry point relative to professional survey-grade equipment. ALPC data from USGS 3DEP is publicly available at no direct cost. For commercial deployments, drone photogrammetry services are typically priced at $150–$300 per acre for standard mapping, with LiDAR-based surveys ranging from $150–$500 per acre depending on point density requirements (The Future 3D, 2026). These figures show that commercial drone survey costs scale substantially with project area, and that municipalities should conduct project-specific cost assessments. Regardless, the framework’s reliance on a consumer-grade drone and publicly available LiDAR data positions it as a cost-accessible alternative to traditional manual inspection campaigns, which are labor-intensive and difficult to scale.

4.7 Future Work

Future work will focus on addressing the framework’s current limitations. We plan to explore more advanced data fusion techniques, such as transformer-based models, to improve accuracy and robustness. Additional sensors, such as thermal imaging for surface quality assessment, will be investigated to enable more comprehensive evaluations. Extending the framework to assess a broader range of pedestrian infrastructure elements,

including sidewalk surface conditions and semantic segmentation, represents a promising direction. Finally, evaluation of the framework's performance across a wider range of urban environments and infrastructure styles will be pursued to establish generalisability.

5. Conclusion

Automated assessment of sidewalk infrastructure for accessibility compliance remains a critical challenge for urban planning agencies seeking to ensure equitable access to public spaces. This paper proposes BirdCV-LiDAR, a multimodal framework that fuses high-resolution bird's-eye-view imagery with aerial LiDAR point cloud data to enable scalable, objective evaluation of sidewalk features against ADA-PROWAG standards. Several conclusions can be drawn from this work: (1) The multimodal data fusion approach achieves 81.0% overall accuracy in compliance determination, significantly outperforming single-modality approaches while maintaining computational efficiency. (2) The framework's reliance on publicly available USGS 3DEP and Natural Resources Canada QL1/QL2 LiDAR data makes it readily deployable across numerous North American urban areas without expensive proprietary data collection. (3) The oriented bounding box detection strategy proves essential for accurate dimensional analysis of rotated sidewalk features, with YOLOv8m-obb demonstrating superior performance despite being architecturally simpler than its successors. (4) Ground-truth validation confirms the framework's practical applicability, with slope assessment accuracy of 84.7% and dimensional RMSE of 0.1152m for crosswalk width measurements.

References

Arefi, H., Hahn, M., Lindenbergh, R., 2020. A review of 3D urban object detection methods from LiDAR point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 160, 1–20.

Armesto, J., Díaz-Vilariño, L., Arias, P., González-Aguilera, D., 2009. Automatic extraction of building facades from mobile LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(2), 150–159.

Axelson, P. W., Chesney, D. A., Galvan, D. V., Kirschbaum, J. B., Longmuir, P. E., Lyons, C., Wong, K. M., 1999. Designing sidewalks and trails for access: Part i of ii: Review of existing guidelines and practices. Technical report, Federal Highway Administration, U.S. Department of Transportation. Prepared by Beneficial Designs, Inc.

Bai, X. et al., 2022. Transfusion: Robust lidar-camera fusion for 3d object detection with transformers. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Balado, J., Martínez-Sánchez, J., Arias, P., Novo, A., 2019. Automatic detection of podotactile paving from aerial lidar data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 19–25.

Balado, J., Martínez-Sánchez, J., Arias, P., Novo, A., 2020. Detection of pedestrian crossings from aerial LiDAR data using orthoimages and spatial reasoning. *ISPRS International Journal of Geo-Information*, 9(2), 121.

Bhat, S., Naik, S., Gaonkar, M., Sawant, P., Aswale, S., Shetgaonkar, P., 2020. A survey on road crack detection techniques. *2020 ic-ETITE*, IEEE.

Board, U. A., 2011. Accessibility guidelines for pedestrian facilities in the public right-of-way.

Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of photogrammetry and remote sensing*, 92, 79–97.

Ding, J., Xue, N., Xia, G.-S., Bai, X., Yang, W., Yang, M., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2021. Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1-1.

Fan, R., Bocus, M. J., Zhu, Y., Jiao, J., Wang, L., Ma, F., Cheng, S., Liu, M., 2019. Road crack detection using deep convolutional neural network and adaptive thresholding. *2019 IEEE IV*, IEEE, 474–479.

Grilli, E., Menna, F., Remondino, F., 2017. A review of point clouds segmentation and classification algorithms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, 339–344.

Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M., Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 93–107.

Hall, D. L., Llinas, J., 1997. An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1), 6–23.

Hassan, S. I., O'Sullivan, D., McKeever, S., Power, D., McGowan, R., Feighan, K., 2022. Detecting patches on road pavement images acquired with 3d laser sensors using object detection and deep learning. *VISIGRAPP (5: VISAPP)*, 413–420.

Himeur, Y., Rimal, B., Tiwary, A., Amira, A., 2022. Using artificial intelligence and data fusion for environmental monitoring: A review and future perspectives. *Information Fusion*, 86, 44–75.

Hoyer, L., 2024. Domain-Robust Network Architectures and Training Strategies for Visual Scene Understanding. PhD thesis, ETH Zurich.

Huber, T., Luecke, K., Hintze, M., Toole, J., Van Oosten, M., 2013. *Guide for Maintaining Pedestrian Facilities for Enhanced Safety', United States*. Federal Highway Administration. Office of Safety, United States.

Ivanchenko, V., Coughlan, J., Shen, H., 2008. Detecting and locating crosswalks using a camera phone. *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, IEEE, 1–8.

Jensen, J. R., Cowen, D. C., 1999. Remote sensing of urban/suburban infras tructure and socio-economic attributes. *Photogrammetric engineering and remote sensing*, 65, 611–622.

Kent, J., 2023. *ADA in details: interpreting the 2010 Americans with disabilities act standards for accessible design*. John Wiley & Sons.

Kyriou, A., Mpelogianni, V., Nikolakopoulos, K., Groumpos, P. P., 2023. Review of Remote Sensing Approaches and Soft Computing for Infrastructure Monitoring. *Geomatics*, 3(3), 367–392.

Li, B., 2018. Application of machine vision technology in geometric dimension measurement of small parts. *EURASIP Journal on Image and Video Processing*, 2018(1), 127.

Li, W., Wei, W., Zhang, L., 2021. GSDet: Object detection in aerial images based on scale reasoning. *IEEE Transactions on Image Processing*, 30, 4599–4609.

Liang, T., Xie, H., Yu, K., Xia, Z., Lin, Z., Wang, Y., Tang, T., Wang, B., Tang, Z., 2022. Bevfusion: A simple and robust lidar-camera fusion framework. *Advances in Neural Information Processing Systems*, 35, 10421–10434.

Natural Resources Canada, 2025. New LiDAR-derived data available on Open Maps! `https://natural-resources.canada.ca/science-data/data-analysis/geospatial-data-tools-services/new-lidar-derived-data-available-open-maps`. Accessed: 2025-10-27.

Nogueira, V. V., Barca, L. F., Pimenta, T. C., 2023. A Cost-Effective Method for Automatically Measuring Mechanical Parts Using Monocular Machine Vision. *Sensors*, 23(13), 5994.

OpenStreetMap contributors, 2017. Planet dump retrieved from https://planet.osm.org . `https://www.openstreetmap.org`.

Pu, S., Rutzinger, M., Vosselman, G., Elberink, S. O., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), S28–S39.

Puente, I., González-Jorge, H., Martínez-Sánchez, J., Arias, P., 2013. Review of mobile mapping and surveying technologies. *Measurement*, 46(7), 2127–2145.

Royimani, L., Mutanga, O., Odindi, J., Dube, T., Matongera, T. N., 2019. Advancements in satellite remote sensing for mapping and monitoring of alien invasive plant species (AIPs). *Physics and Chemistry of the Earth, Parts A/B/C*, 112, 237–245.

Sadik-Khan, J., 2012. Urban street design guide. *New York: NACTO*, 554.

Schofield, O. B., Iversen, N., Ebeid, E., 2022. Autonomous power line detection and tracking system using UAVs. *Microprocessors and Microsystems*, 94, 104609.

Smith, V., Malik, J., Culler, D., 2013. Classification of sidewalks in street view images. *2013 International Green Computing Conference Proceedings*, IEEE, 1–6.

The Future 3D, 2026. Drone Survey Cost Guide: Pricing for Aerial Mapping in 2026. `https://www.thefuture3d.com/learn/drone-survey-cost-guide/`. Accessed: 2026-03-30.

US Department of Justice, 2015. *AMERICANS WITH DISABILITIES ACT TITLE II REGULATIONS*. EISENBRAUNS.

U.S. Geological Survey, 2025. 3D Elevation Program (3DEP). `https://www.usgs.gov/3d-elevation-program`. Accessed: 2023-10-13.

Wang, H., Liu, J., Dong, H., Shao, Z., 2025. A survey of the multi-sensor fusion object detection task in autonomous driving. *Sensors*, 25(9), 2794.

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

Williams, J. E., Evans, M., Kirtland, K. A., Cavnar, M. M., Sharpe, P. A., Neet, M. J., Cook, A., 2005. Development and use of a tool for assessing sidewalk maintenance as an environmental support of physical activity. *Health promotion practice*, 6(1), 81–88.

Wu, Y., Liu, Q., Sun, H., Xue, D., 2024. HRTracker: Multi-Object Tracking in Satellite Video Enhanced by High-Resolution Feature Fusion and an Adaptive Data Association. *Remote Sensing*, 16(17), 3347.

Zhang, W., Ning, Y., Suo, C., 2019. A method based on multi-sensor data fusion for UAV safety distance diagnosis. *Electronics*, 8(12), 1467.