INCORPORATING USE INSPIRED DESIGN IN PROVIDING SAFE TRANSPORTATION INFRASTRUCTURE FOR RITI COMMUNITIES

FINAL PROJECT REPORT
by

Roger B. Chen[1], Poya Harirchi, Qi Chen[2], Adrian Archilla[2], Preston Garcia, Xiazhi Zhang and Shiva Azimi

[1]Principle Investigator

[2]Co-Principle Investigator

Department of Civil, Environmental and Construction Engineering
University of Hawaii at Manoa

Center for Safety Equity in Transportation (CSET)
USDOT Tier 1 University Transportation Center
University of Alaska Fairbanks
ELIF Suite 240, 1764 Tanana Drive
Fairbanks, AK 99775-5910

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The Center for Safety Equity in Transportation, the U.S. Government and matching sponsor assume no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
INCORPORATING USE INSPIRED DESIGN IN PROVIDING SAFE TRANSPORTATION INFRASTRUCTURE FOR RITI COMMUNITIES	6/30/2024
	6. Performing Organization Code

7. Author(s) and Affiliations	8. Performing Organization Report No.
Roger B. Chen, Poya Harirchi, Qi Chen, Adrian Archilla, Preston Garcia, Xiazhi Zhang and Shiva Azimi – UH-Manoa, Department of Civil, Environmental and Construction Engineering	INE/CSET 24.15

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Center for Safety Equity in Transportation ELIF Building Room 240, 1760 Tanana Drive Fairbanks, AK 99775-5910	
	11. Contract or Grant No.
	69A3551747129

12. Sponsoring Organization Name and Address	13. Type of Report and Period Covered
United States Department of Transportation Research and Innovative Technology Administration 1200 New Jersey Avenue, SE Washington, DC 20590	Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Report uploaded to:

16. Abstract

In this study, we focus on automating road marking extraction from the HDOT MLS point cloud database, managed by Mandli. Mandli is a company specializing in highway data collection, including LiDAR. Mandli has cooperated with various Department of Transportation throughout the United States. Here, we focus on infrastructure elements related to non-motorized travel modes, supporting the ongoing Complete Streets efforts in Hawaii. Point cloud data include different colors that represent differences in elevation and intensity values. Based on a visual inspection, road markings can be observed within these point clouds. The long-term objective of this study is to develop a framework and approach for automating the detection of these infrastructure elements, based on deep learning approaches. For this project, a YOLOv5 (You Only Look Once version 5) image object detection model was trained with the HDOT point cloud data. YOLO is a family of deep learning models designed for fast object detection; the latest published version is the 5th version. The focus here is on non-motorized objects, such as crosswalks, bike lanes and bike boxes. The same approach can be extended to other markings as well, which we plan for subsequent studies.

17. Key Words	18. Distribution Statement
Safety, Pedestrian, LiDAR	

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified.	Unclassified.	41	N/A

Form DOT F 1700.7 (8-72)   **Reproduction of completed page authorized.**

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

In this project, the research team investigated the feasibility of LiDAR data for extracting infrastructure geometries and other human scale features relevant for pedestrian safety from LiDAR data collected annually for roads managed and maintained by the Hawaii State Department of Transportation (HDOT), and their potential for informing policies that provide safe accommodations for all travelers. LiDAR data has been widely used in transportation for mapping, object detection, and capturing highly accurate and detailed data. Our research initially focused on road markings since they serve as fundamental visual cues on roads and highways, providing crucial information, guidance, and regulation for drivers, pedestrians, and cyclists. The current absence of an automated method for recognition of road markings has resulted in time-consuming and labor-intensive processes as well as inaccurate and error-prone results. Developing automated methods for extraction, detection, and localization of road markings can overcome these challenges and enable more efficient, accurate, and up-to-date road marking maintenance management, benefiting both road users and transportation authorities. This investigation into road markings lays the groundwork for future explorations, utilizing LiDAR data to uncover a broader range of signs and signals embedded within the transportation landscape.

The Hawaii State Department of Transportation (HDOT) has periodically collected photologs of their roadways since 2003, and in 2009 began also collecting LiDAR for the island of Oahu, with the islands of Hawaii (the Big Island), and Maui/Kauai following in 2011 and 2012 respectively. Given the long-term objective of this study to develop a framework and approach for automating the detection of infrastructure elements based on deep learning approaches, a YOLOv5 (You Only Look Once version 5) image object detection model was trained with the HDOT point cloud data. Our model was able to accurately detect 85% of road-markings in the dataset, with results closer to 92% for crosswalks. These results show that the automatic detection and geocoding of relevant roadway assets may be possible through the analysis of LiDAR point cloud data using the framework and methods utilized in this study.

Utilizing LiDAR point cloud data for detection of road markings also has certain advantages compared to satellite imagery. Examples of these advantages include detection of markings that would otherwise be blocked by other road facilities and natural obstructions such as trees. Other advantages of this framework include the elimination of the need for inspections and related work zone safety issues, and permitting quicker detection and response to changing road conditions.

Finally, a Poisson Regression Analysis was performed on the City and County of Honolulu crash counts to analyze safety across Census tracts, and to determine the relationship between detection and confidence of detection (as a measure of quality) of roadway markings and pedestrian and bike crash incidents. The model shows that Census tracts with a greater detection of crosswalks showed higher crash rates, and tracts with a higher mean confidence scores (related to marking quality) showed lower crash rates. Similar to results for pedestrian crashes, the model shows that tracts with a greater detection of bike symbols showed high crash rates, and tracts with detected symbols that have a higher mean confidence (greater marking quality) showed lower crash rates. Furthermore, it can be concluded that higher population density tracts are associated with lower crash rates, and higher roadway length to tract area ratios are associated with higher crash rates, though this was statistically insignificant for bike crashes. For pedestrian crashes, tract areas with higher areas of Hawaiian Homelands per tract area show lower pedestrian crash rates, but this was not too statistically significant.

CHAPTER 1.   INTRODUCTION

LiDAR (short for Light Detection and Ranging) point cloud sensors, are found in a wide range of technologies, including airplanes, autonomous vehicles, and consumer smartphones. LiDAR sensors operate by emitting pulsed light waves into the surrounding environment. When the light pulses bounce back to the sensor, the sensor tracks the time it took for each pulse to return to the sensor, and the angle of pulses received to calculate the 3-dimentional (3D) coordinates for each LiDAR point. Repeating this process millions of times produces a point cloud, which is a cluster of points in 3D space. Each point in the point cloud contains multiple pieces of information including spatial coordinates, intensity values, RGB values, and bounce angles.

LiDAR is widely used in transportation for mapping, object detection, and capturing highly accurate and detailed data. Within the wide array of transportation data captured by LiDAR point clouds, our research initially focused on road markings. Road markings serve as fundamental visual cues on roads and highways, providing crucial information, guidance, and regulation for drivers, pedestrians, and cyclists. This investigation into road markings lays the groundwork for future explorations, utilizing LiDAR data to uncover a broader range of signs and signals embedded within the transportation landscape.

Using the information attributed to the points and their spatial geometries as inputs, it is possible to detect objects in point clouds through deep learning approaches. Compared to 2D photo images, point clouds have multiple advantages. Accurately extracting information regarding an object's actual dimensions is difficult because 2D images lack depth information. Each point of a point cloud is geocoded and contains 3D information, creating a potential for inferring objects' geometries and the spatial relationships between objects. Furthermore, LiDAR does not require external light to collect data like conventional cameras (1). LiDAR is less affected by variations in lighting conditions compared to images and videos. LiDAR operates based on physical properties like geometry and reflectivity, rather than relying on visual appearance like images and videos. LiDAR point cloud data allows for efficient filtering and removal of other objects present in the scene, such as vehicles or pedestrians. This capability reduces false detections caused by objects that may resemble road markings in images or videos. LiDAR point cloud overcomes visual barriers caused by objects like vegetation and overpasses. This capability provides unobstructed access to the road surface and enhances the reliability and completeness of road marking analysis.

Mobile LiDAR Systems (MLS) are used widely to gather infrastructure information, where the LiDAR sensor is mounted on the roof of a vehicle and collects data as the vehicle is moving. One advantage of MLS is the ability to collect data while moving in traffic, minimizing the disruption to traffic and shortening collection times. The information collected has supported analyzing road markings, pavement condition, road edge information, traffic signs, traffic lights, etc. since the sensor covers a 360-degree range. The benefits from MLS have resulted in several State DOTs initiating efforts to collect point clouds of their roadways, including Oregon Department of Transportation and Massachusetts Department of Transportation (1). The Hawaii State Department of Transportation (HDOT) has collected photologs of their roadways since 2003, and in 2009 began also collecting LiDAR for the island of Oahu, with the islands of Hawaii (the Big Island), and Maui/Kauai following in 2011 and 2012 respectively (2). Alongside this data collection effort, the City and County of Honolulu and other parts of the state have begun embracing the Complete Streets Design perspectives, which promotes more use of non-motorized and public transit travel modes (3). These developments underscore the importance of point

cloud information for supporting asset management, specifically for non-motorized travel modes. As LiDAR point clouds continues to increase in frequency and volume, an opportunity exists for extracting information from them that is pertinent to the Complete Streets Design efforts.

The absence of an automated method for recognition of road markings results in time-consuming and labor-intensive processes as well as inaccurate and error-prone results. Road markings are subject to damage and wear due to factors like traffic load and weather conditions. To maintain effective road marking visibility and functionality, regular maintenance and timely reapplication of markings are necessary. In the absence of an automated method, updating and reflecting these changes in a timely manner proves challenging. Developing automated methods for extraction, detection, and localization of road markings can overcome these challenges and enable more efficient, accurate, and up-to-date road marking maintenance management, benefiting both road users and transportation authorities.

In this study, we focus on automating road marking extraction from the HDOT MLS point cloud database, managed by Mandli. Mandli is a company specializing in highway data collection, including LiDAR. Mandli has cooperated with various Departments of Transportation throughout the United States (4). Here, we focus on infrastructure elements related to non-motorized travel modes, supporting the ongoing Complete Streets efforts in Hawaii. Figure 1 presents an example of point clouds for a street in Honolulu. Different colors represent differences in elevation and intensity values. Based on a visual inspection, road markings can be observed within these point clouds. The long-term objective of this study is to develop a framework and approach for automating the detection of these infrastructure elements, based on deep learning approaches. For this project, a YOLOv5 (You Only Look Once version 5) image object detection model was trained with the HDOT point cloud data. YOLO is a family of deep learning models designed for fast object detection; the latest published version is the 5th version (5). The focus here is on non-motorized objects, such as crosswalks, bike lanes and bike boxes. The same approach can be extended to other markings as well, which we plan for subsequent studies.

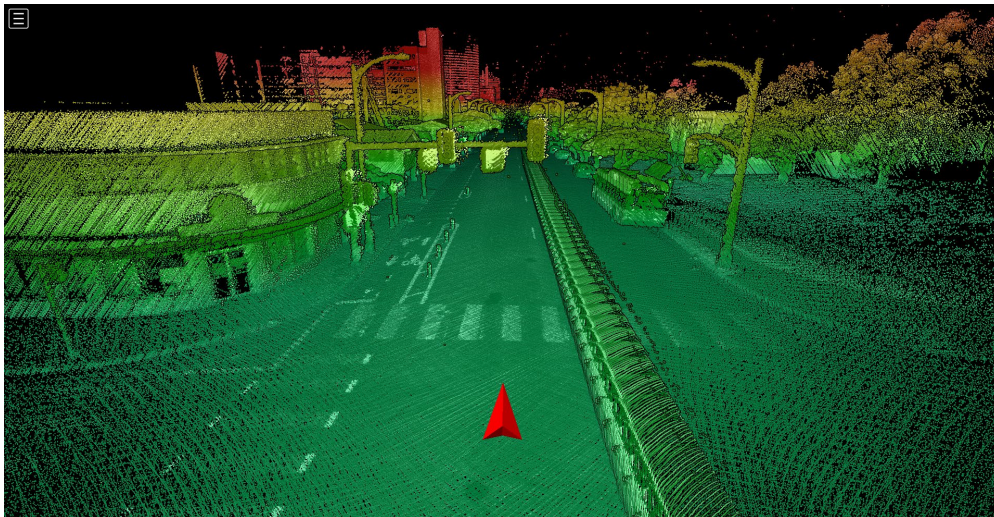Figure 1. Example of LiDAR point clouds in Honolulu (Isenberg Street and King Street intersection)

Finally, a Poisson Regression Analysis was performed on the City and County of Honolulu crash counts to analyze safety across Census tracts, to determine the relationship between detection and

confidence of detection (as a measure of quality) of roadway markings and pedestrian and bike crash incidents, as discussed in Chapter 4.

CHAPTER 2. REVIEW OF LITERATURE

Road markings play a crucial role in high-precision maps, which are increasingly employed to manage and regulate traffic activities in intelligent transportation, travel behavior, and autonomous vehicles (6, 7). It is vital to consider the influence of crosswalks and pedestrians' behavior when analyzing pedestrian crash records since these locations serve as points of interaction between pedestrians and vehicles. Previous studies have shown that approximately 90 percent of pedestrian-vehicle collisions occurred within 70 feet of a crosswalk (8). Detecting the precise locations of crosswalks is therefore highly beneficial in identifying and preventing such incidents in the future, particularly with the rise of autonomous vehicles. Studies on crosswalk detection and analysis from images emerged in the late 20th century. The first investigation into road surface analysis was conducted by Pomerleau (9) using the Artificial Neural Network-based Autonomous Land Vehicle (ALVINN) in 1989. Kitawaki et al. (10) attempted to estimate the length of crosswalks by converting real images to grayscale to assist visually impaired individuals in crossing independently. Se et al. (11) employed simple computer vision techniques such as the Hough transform and Canny edge detector to detect crosswalks in images.

The advent of Deep Learning (12) and Convolutional Neural Networks (CNNs) revolutionized object detection in images. Deep learning methods have been utilized for detecting objects, pedestrians, and traffic signs on highways (13, 14, 15). Malbog (16) employed Mask R-CNN, which utilizes ResNet as its backbone network, to detect crosswalks from dash camera images and internet videos, achieving an accuracy of 97% for the model. The introduction of You Only Look Once (YOLO), a unified and real-time object detection algorithm, has further transformed the field, constantly improving speed and precision (17, 18, 19). YOLO has gained popularity in the object detection community, as demonstrated by Zhang et al. (20), who developed a Convolutional Neural Network based on YOLOv5 called CDNet. This network achieved an F1 score above 94% in detecting crosswalks from real-time images. It is important to note that factors such as weather and illumination can significantly impact crosswalk detection from images. Moreover, a lack of image preprocessing leads to inconsistent results, affecting object classification and detection (21). In this regard, the use of Laser Imaging Detection and Ranging (LiDAR) proves advantageous, as road markings can be detected using LiDAR point clouds even without external illumination. Hata et al. utilized LiDAR reflective intensity data to detect road markings (22), and other studies have employed LiDAR systems to determine intersections on roads (23, 24, 25). Previous studies have projected (3D) point clouds onto 2D images to detect crosswalks from images, but the question of crosswalk geolocation remains unanswered. This study aims to address both crosswalk detection and geolocation.

In past studies, (3D) point clouds are first projected onto 2D images. Taking advantage of reflectivity difference between asphalt and road markings, a segmentation is performed to segment road markings from other objects in the point cloud. A pavement segmentation is sometimes used prior to segmentation to eliminate all the points above the ground and greatly reduce the computational resource required. Soilan et al. (26) propose a method based on the construction of a saliency map for pavement segmentation of point clouds. The point clouds are first partitioned into sections with equal length, in the direction of MLS vehicle's trajectory. Second, the dominant normal vector of cloud is computed using K-means clustering. Finally, the distance between each normal vector and the dominant normal vector is projected into a hyperbolic tangent function apace, making the difference between a salient and non-salient point is large or salient enough for classification. Yao et al. (27) propose a

method based on the distance between scan lines for pavement segmentation and develop thresholds and conditions for distance leap between adjacent points to extract scanlines. Finally, the geometric features, such as elevation and slope, are calculated for each scanline to extract road surface. Chen et al. (28) convert point clouds into intensity images by using inverse distance weighted (IDW) interpolation to determine the intensity value for each grid. Wen et al. (29) and Jung et al. (1) projected 3D point clouds onto gridded 2D image; the average density of each cell was used as the density value.

For road marking segmentation, past studies have used the intensity value as the main attribute. Soilan et al. (26) propose that the intensity distribution of roads with markings can be further distributed into two classes that approximate Gaussian distributions. The segmentation follows from properly classifying the points, and then applying a Gaussian Mixture Model with two clusters to filter the point clouds. The same authors use a normalized sum of intensities as the feature for creating the intensity-based image. Finally, an adaptive thresholding approach was used to improve the masking performance. Yao et al. (27) took a similar direction to Soilan et al. (26), which included point cloud rasterization, binary segmentation, and a noise filtering process. Yao et al. (27) use IDW interpolation to generate intensity image, then marking segmentation was performed by using adaptive threshold that based on an integral image. Wen et al. (29) and Chen et al. (28) took different approaches to perform extraction; they utilized machine learning based methods to extract features automatically. Wen et al. (29) constructed a modified U-net to perform road marking classification at the pixel level. Images with different intensity were used in training to solve the intensity variation problem. U-net is a convolutional neural network (CNN) originally developed for biomedical image segmentation; it is widely used as a semantic segmentation method now. Chen et al. (28) proposed a dense feature pyramid network-based deep learning model for road marking segmentation. A U-net network was adopted for feature extraction and feature pyramid networks were established with a residual neural network (ResNet). ResNet is a type of neural network widely used in computer vision that allows for training for extremely deep neural networks (30). Next, a region proposal network was established for calculating candidate regions; these region proposals were mapped into feature pyramid networks to generate regions of interests. They were used in a ROI predictor network to complete the detection and segmentation. The next step generates semantic information for marking segmentation.

Soilan et al. (26) propose a machine learning algorithm with geometry-based features to perform semantic classification on road markings; the neural network would first classify markings into three generic classes. Then the geometric features in each class were used to assign a proper semantic meaning to the road markings. Yao et al. (27) and Wen et al. (29) used geometric features for semantic classification of larger road makings mostly lane lines and crosswalks. The geometric features were collected through Euclidean distance clustering and minimum bounding rectangle. Yao et al. (27) used a template matching method for arrow markings; the skeleton of arrow markings was extracted and matched with road marking design standards. Wen et al. (29) trained a convolutional neural network classifier for identifying road markings other than lane lines and crosswalks.

6

CHAPTER 3.   DATA AND METHODS

3.1.   Stage 1: Pilot Study

In the first stage of this project, we investigated the feasibility of a framework and approach for extracting road marking information from LiDAR point clouds collected by MLS systems used by the Hawaii State DOT, on a subset of targeted streets in Honolulu. In this framework, the LiDAR point cloud was projected onto a 2D intensity image. A deep learning-based object detection model was trained and applied to detect road markings, given this 2D intensity image. A model from this approach achieved a 0.85 average precision at an Intersection over Union (IoU) threshold equal to 0.50.

Based on the review of approaches for extracting markings from point clouds of roadways, we proposed the following framework and approach.

Figure 2. Flowchart for the Workflow
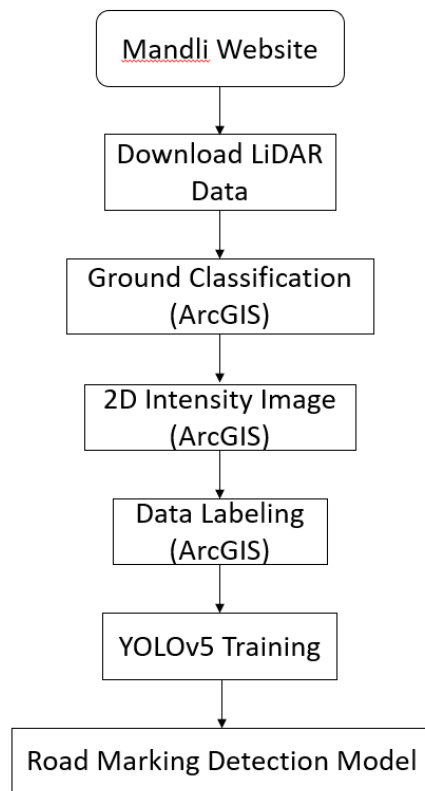
While some parts overlap with previous studies, our approach is convenient and is intended for engineering at HDOT. Figure 2 presents the workflow of our approach. The approach can be completed primarily by using ArcGIS, which is widely available. This approach provides another option for engineers without proficient knowledge in computer vision technology and implementation.

3.1.1. Data Acquiring and Preprocessing

Before any analysis can be performed on the point clouds, these data need to be extracted from the database of LiDAR data collected. This section discusses the data acquisition process. The point cloud data for the targeted streets in Honolulu were downloaded. The data only has 4 attributes; they are the x-y-z coordinates and the intensity values. Intensity is the main attribute used for road marking detection in this study. Crosswalk markings and bike sharrows are the two types of road markings considered in this study. To acquire enough samples, point clouds for ten roads were selected for downloading; nine of them were for training and one was kept for validation. These ten roads were selected based on the discussions with HDOT and the City and County of Honolulu (CCH) Department of Transportation Services (DTS). All of these are major arterials in Honolulu with a broad combination of traffic markings. The roads selected are listed in Table 1 and shown in Figure 3.

Table 1. Training and validation road selection

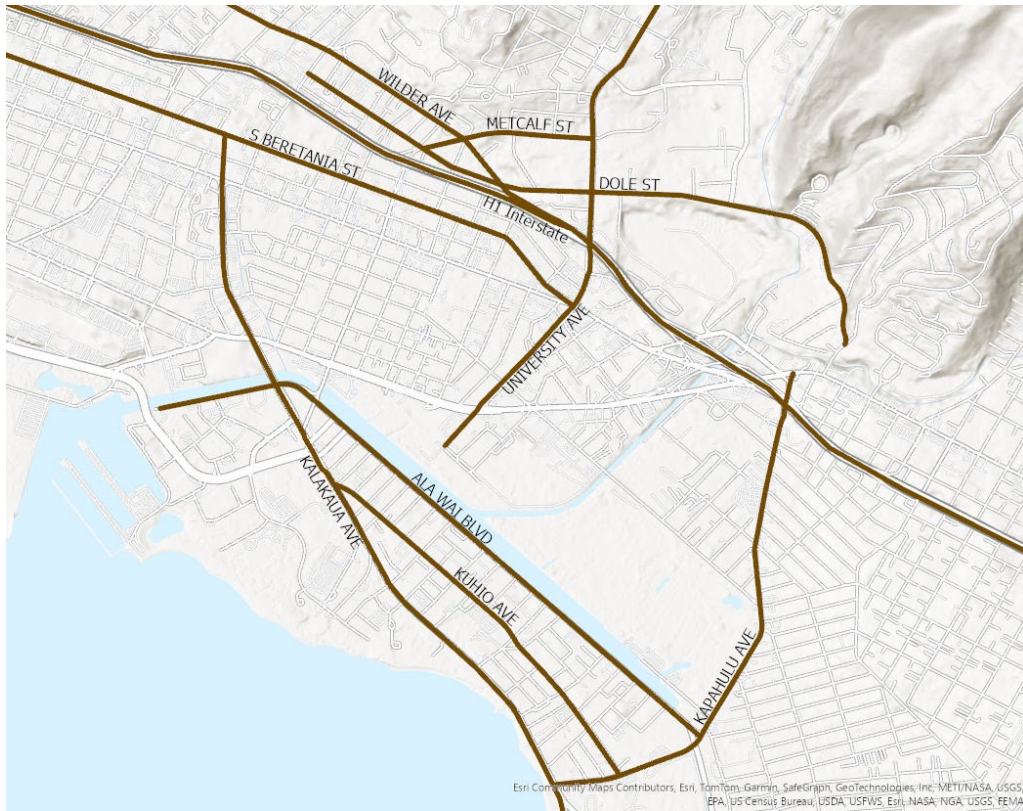| Street Name	Purpose	Street Name	Purpose
Dole Street	Training	University Avenue	Training
Metcalf Street	Training	Wilder Avenue	Training
Beretania Street	Training	King Street	Training
Ala Wai Blvd	Training	Kuhio Avenue	Training
Kalakaua Avenue	Training	Kapahulu Avenue	Validation

Figure 3. Map of selected roads for training and validation

Dole Street, University Avenue and Metcalf Street are the roads around the University of Hawaii at Manoa. University Avenue is one of the major roads that provides access to Manoa valley. Metcalf Street and Wilder Avenue are the major roads north of the H1 Interstate. The traffic exiting H1 use these streets to access local areas, and are thus, extremely high volume. Beretania Street and King Street are the two major east-west streets for Honolulu. Ala Wai Boulevard, Kalakaua Avenue and Kuhio Avenue are three major roads in Waikiki, which is most popular destination for tourists in Honolulu. Kapahulu Avenue is a major road that starts in Waikiki and travels north with a heavy concentration of businesses along its side. The roads selected are all major high-volume roads and have many intersections.

3.1.2. *Obtaining Samples and Drawing Bounding Boxes*

Based on the objects of interest to this study, there are two approaches for extracting point clouds. The first approach considers intersections. Most of the intersections will have at least one crosswalk markings. We draw a single bounding box for one intersection. The boundary of this box extends beyond the edge of the crosswalk markings for at least the width of the crosswalk markings in all directions. In some cases, there will be bike sharrows that are close to the crosswalks, which are included within the bounding box. The second approach considers the middle of roads, away from intersections. Bike sharrows and midblock crosswalks are typically present. In this case, we draw a bounding box to enclose that single object. If there is another target object nearby, the bounding box should be including both objects.

Before training the deep learning model, data preprocessing and labeling were conducted on the downloaded data. The data labeling and preprocessing process are all completed with ArcGIS Pro. The data preprocessing procedure is presented as the following:

1. The points above the ground were all removed using the ArcGIS ground identification tool.
2. The remaining ground points were projected onto horizontal plane and then rasterized with a resolution based on the averaging point spacing.
3. The average intensity value for points projected onto each pixel was assigned a gray value. In this study, 0.04m was used as resolution and the *natural neighbor interpolation method* was used to fill voids.
4. The road markings were labeled on the intensity images within ArcGIS.
5. The labeled intensity images were exported to uni-sized images to use as the input data for model training.

There are 237 crosswalk markings and 86 bike sharrows labeled in training set; 48 crosswalk markings and 31 bike sharrows labeled in the validation set. Figure 3 shows the visualization of intensity image and point cloud. Figure 4 shows the samples of exported training images with bounding boxes and class labels.
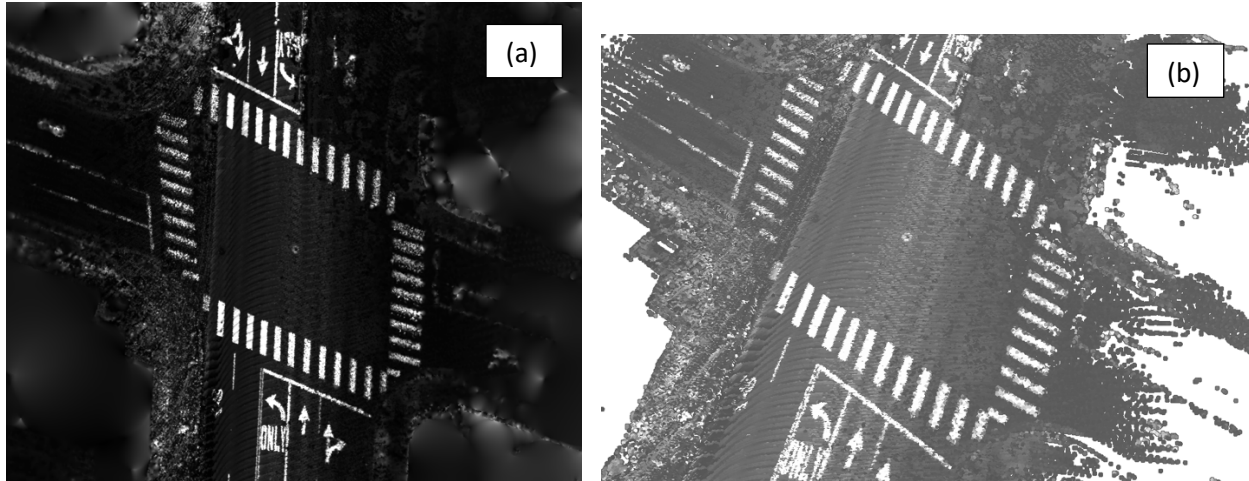
Figure 4. The intensity image and point cloud at the same location. (a) intensity image (b) point cloud.
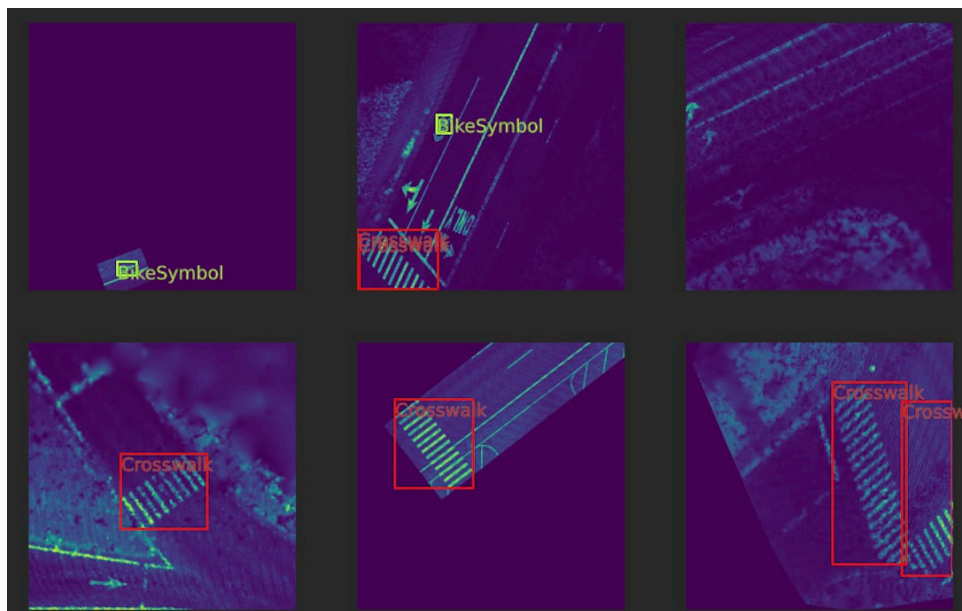
Figure 5. Exported fixed size training image samples

3.1.3. Training Deep Learning Model

We trained a YOLOv5m (You Only Look Once version 5 median size model) model for road marking detection and classification. YOLO is an algorithm that uses neural networks to provide read-time object detection. Detection and classification are two different tasks. In detection the model is trained to detect the possible location and size of the object on image; in classification the model is trained to recognize the type of objects only in the detected region from detection task. The model was trained on a desktop with a 6th generation i7 CPU with 16 GB memory. The model was trained for 90 epochs, each

epoch takes about 30 minutes. One epoch is when the training process has gone through all the samples completely.

3.1.4. Findings and Results

This section presents the findings and results of the pilot study. The metrics considered for training the model include: (i) training loss; (ii) validation loss; (iii) recall; (iv) precision; and and (v) mAP. The loss function is composed by 3 parts: box loss, *objectness loss*, and *classification loss*. The box loss is loss for bounding box regression. The *objectness loss* is the confidence of an object presence in the anchor, when a bounding box is predicted, there is a confidence score based on whether object is actually presented in the predicted bounding box, the *objectness loss* is based on this. The *classification loss* is the loss for object classification (*5*). A metric commonly used for object detection is *intersection over union* (IoU). In object detection, there is a ground truth bounding box and predicted bounding box. If the predicted box is the same dimension and completely overlaps the ground truth box, then we have a perfect detection. The IoU calculates the ratio of the area of overlap to the area of union. A higher IoU value indicates the predicted the bounding box closely resembles the ground truth bounding box. A threshold of IoU is defined in detection training. If IoU value is above the threshold, then it is considered as a successful detection. In this study the IoU threshold was set to 0.5. In machine learning classification problems, a confusion matrix, also known as an error matrix, is used for summarizing the performance of a classification model. The rows in confusion matrix represent the instance in an actual class, while the columns represent instances as a predicted class. Table 2 shows the basic structure of a confusion matrix. True positive (TP) means the sample is positive and the model successfully predicts this sample to be positive. False negative (FN) means the sample is positive but the model falsely predicts this sample to be negative. False positive (FP) means the sample is negative but the model falsely predicts this sample to be positive. True negative (TN) means the sample is negative and the model successfully predicts this sample to be negative. The calculation for other performance measures is based on the values in the confusion matrix.

Table 2. Confusion Matrix

		Prediction Condition	
	Total Population = P+N	Predict Positive (PP)	Predict Negative (PN)
Actual Condition	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Positive (FP)	True Negative (TN)

Accuracy is the ratio of number of correct predictions to total number of observations. It is the most intuitive performance measure. But when the dataset is unbalanced, accuracy may not be a good measure for model performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \ (Eq.\,1)$$

Recall is the ratio of correctly predicted positive observations to all observations in actual class. It measures the ability of a model to find all the positive cases within a dataset. A high recall value means the model is performing well in identifying all true positive cases. Recall is a good performance metric when there is high cost associated with false negative. But recall cannot tell if there are true negative cases identified as false positive cases.

11

$$Recall = \frac{TP}{TP + FN} \ (Eq. 2)$$

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. It measures the ability of a model to find only the positive cases. A high precision value means most of the positive predictions are true positive cases. Precision is a good performance metric when the cost of false positive is high. Precision cannot tell if there are true positive cases identified as false negative cases.

$$Precision = \frac{TP}{TP + FP} \ (Eq. 3)$$

Mean average precision (mAP) is the average precision for all recall values and for all classes at current IoU threshold. It is also the area under precision-recall curve. It is a metric that combines both recall and precision.

Figure 5 shows the loss curves for training and validation. The curves convergence after 80 epochs; no sharp zigzags and no obvious overfitting can be observed. Table 3 presents the training results in detail. The *mAP_0.5* refers to the mAP with an IoU threshold of 0.5. The *mAP_0.5:0.95* refers to the value for the average of all mAP values with IoU threshold from 0.5 to 0.95 with 0.05 incremental interval. The trained model reaches a *mAP_0.5* value at about 0.85, which means if IoU threshold is 0.5, the mean average precision is about 0.85. Figures 6 and 7 shows some examples for the comparison between ground truth and predicted results. From this batch of samples, the trained model can successfully identify most of the crosswalks.

Table 3. Training Results

| Epoch	Train loss	Val loss	Precision	Recall	mAP_0.5	mAP_0.5:0.95
1	0.154237	0.137535	0.0079357	0.0086634	0.0017359	0.00058931
10	0.0931694	0.084464	0.51194	0.39504	0.37174	0.15406
20	0.0722603	0.0613809	0.70249	0.56879	0.63229	0.31862
30	0.0597233	0.0520671	0.87268	0.65149	0.73951	0.47735
40	0.0643798	0.0636455	0.63629	0.54802	0.54992	0.2677
50	0.0575146	0.0528312	0.92266	0.66759	0.76652	0.47973
60	0.0496196	0.0485755	0.92898	0.6928	0.79445	0.52391
70	0.0456733	0.0410717	0.94102	0.72999	0.8232	0.60152
80	0.0420052	0.0393135	0.9667	0.75032	0.83804	0.62662
90	0.0393014	0.0364516	0.96957	0.75727	0.84943	0.66525
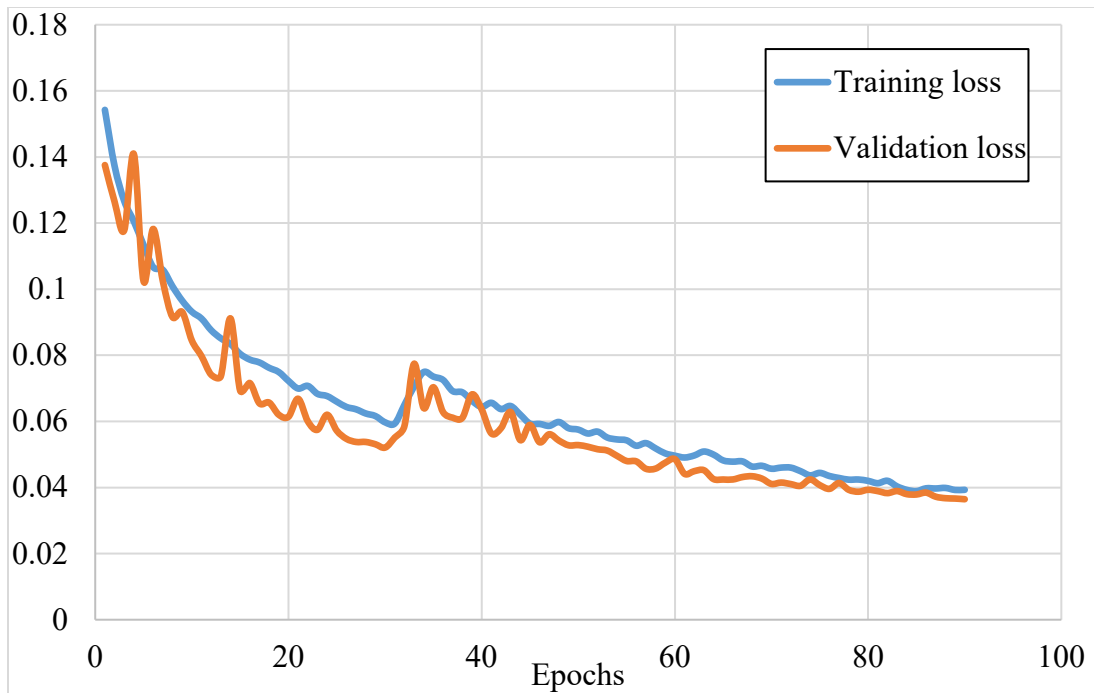
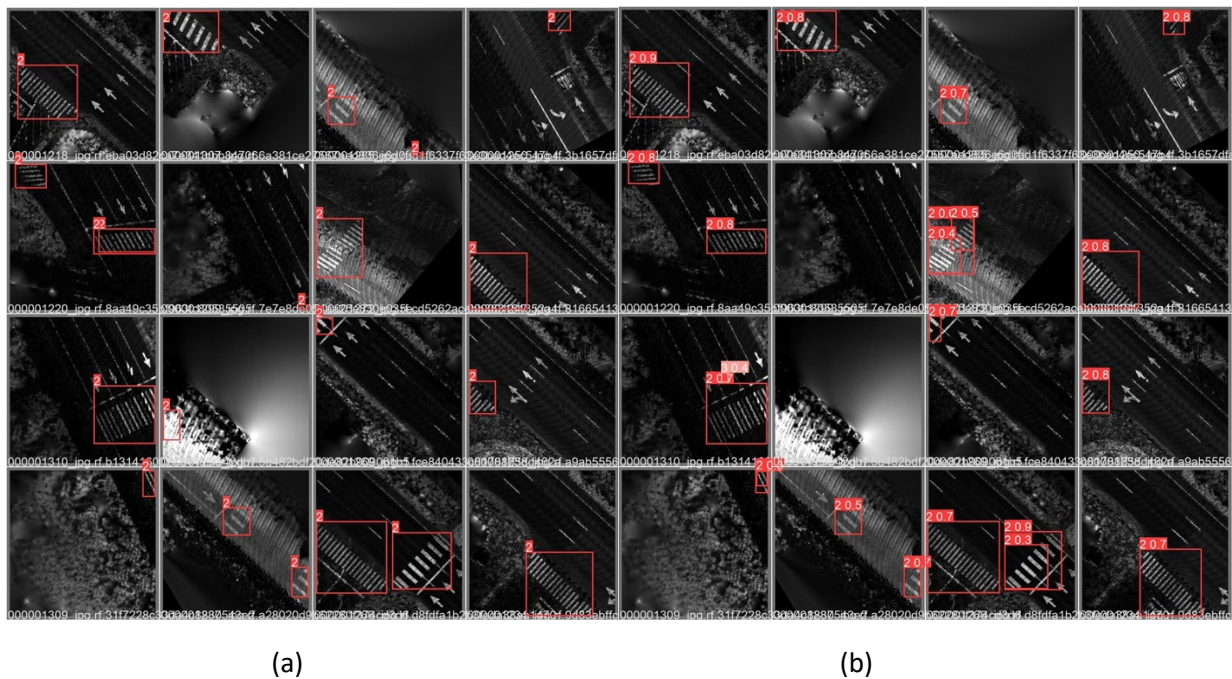Figure 6. Training and validation loss curves

| (a) | (b) |

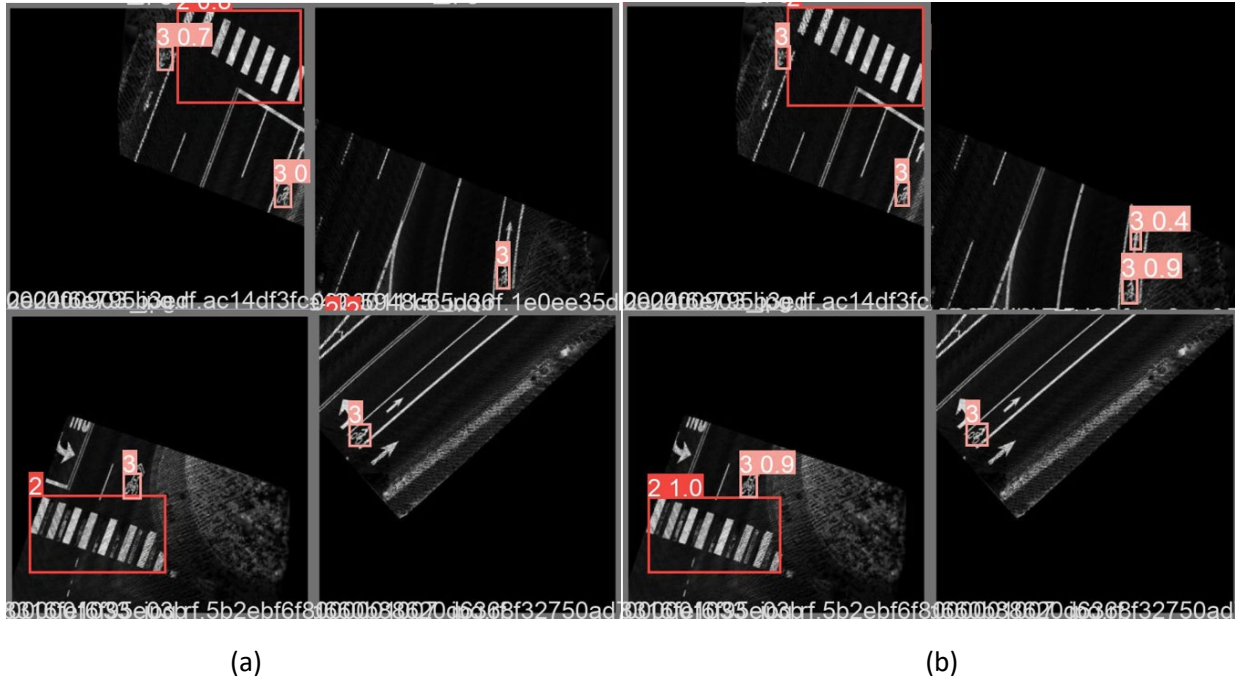Figure 7. Examples of ground truth and predicted results for crosswalks, (a) ground truth, (b) predicted results.

Figure 8. Examples of ground truth and predicted results for bike sharrows, (a) ground truth, (b) predicted results.

3.2. Stage 2: Automatic Road Marking Detection and Geo-location From LiDAR Point Clouds for City and County of Honolulu Roads

For the second stage of this project, after the feasibility of a framework and approach for extracting road marking information from LiDAR point clouds collected by MLS systems used by the Hawaii State DOT was established on a subset of streets as described in section 3.1., the study area was expanded to include the entire area covered by the City and County of Honolulu.

3.2.1. Data Acquisition and Processing

The data acquisition and processing for Stage 2 of this project included the following steps:

Step 1: The data processing begins with acquiring LiDAR points of road segments collected by Mandli, accessed through the Mandli Web Portal for HDOT ([21]). We specifically targeted road segments with roadway markings, specifically crosswalks and bike symbols. These segments were then processed to extract the desired road markings, which were subsequently enclosed within 3D bounding boxes. A representative instance is showcased in Figure 8.
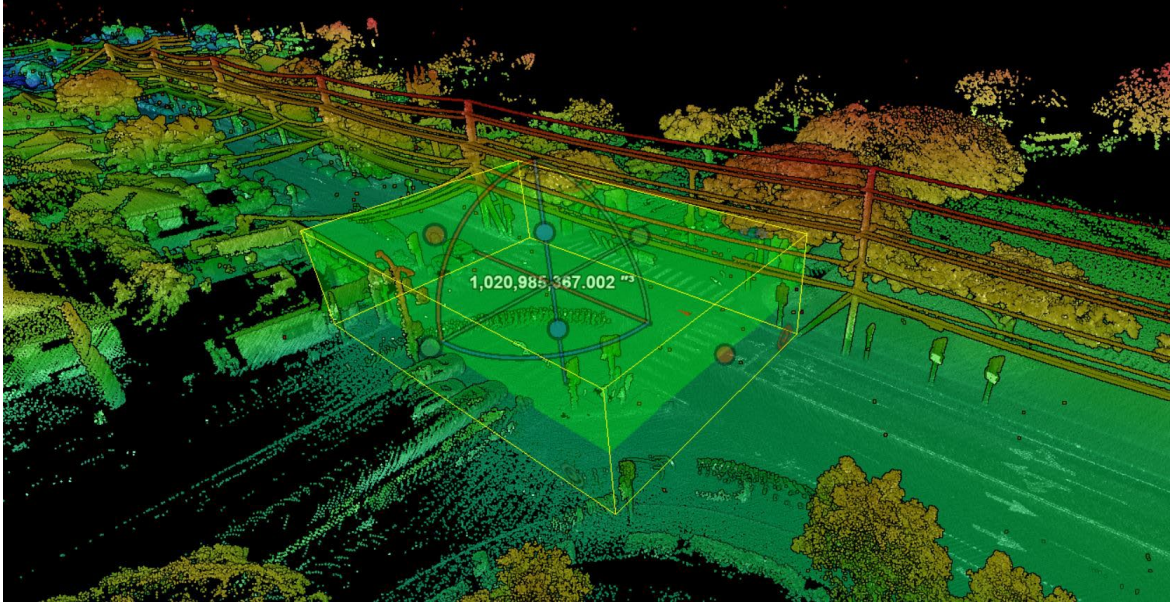
Figure 9. Example of a captured bounding box around a road segment

Selecting road segments manually only based on the presence of road markings can be a time-consuming process. An alternative approach is automating the selection process by systematically moving along the road and exporting all encountered segments. By including a wide range of heterogeneity in the road markings, we can develop a better model to detect road markings, even those outside our specific focus. This expanded dataset enhances the model's accuracy in identifying the road markings of interest. The dataset at this stage comprises *LAS* files, which is a standard file format for the interchange and archiving point cloud data. Each LAS file contains a road segment's point cloud. Our next step is to extract the ground points from point cloud data.

Step 2: We performed a point cloud classification to differentiate between ground and non-ground points. By filtering out non-ground points, we narrow our focus to the road surface, improving the accuracy of road marking detection. Refining the road surface by filtering out non-ground points eliminates potential obstacles such as overhead electric light wires, ensuring a clearer representation of the road markings (Figure 9). By utilizing the 3D nature of the point cloud data, we can achieve a more unobstructed view of the road surface.
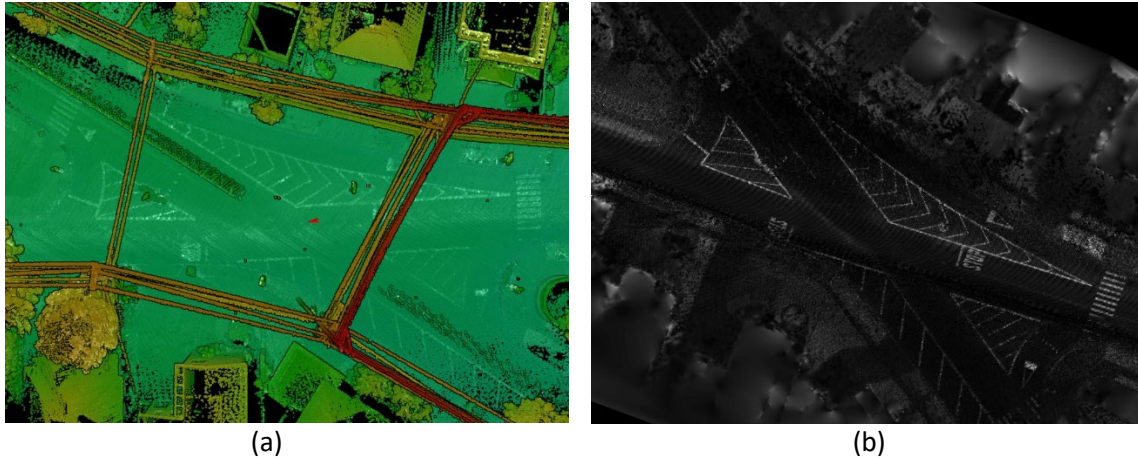
(a)　　　　　　　　　　　　　　　　　(b)

Figure 10. Removing Non-Ground Points: a) Point Cloud with Non-Ground Points (including overhead electric light wires); b) 2D Intensity Image after Removing Non-Ground Points

Step 3: After removing non-ground points, we convert each point cloud into a 2D intensity images by projecting the three-dimensional point cloud data onto a two-dimensional plane. Figure 10 exemplifies this transformation.
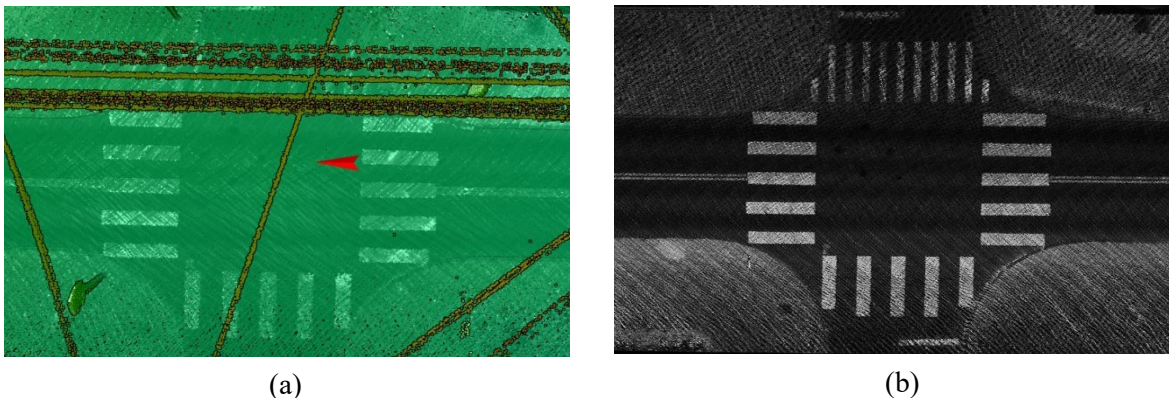

(a)　　　　　　　　　　　　　　　　　(b)

Figure 11. (a) original LiDAR data, and (b) corresponding 2D intensity images from the same point cloud

2D images provide a more intuitive and easily understandable representation of the 3D LiDAR data. Converting point clouds to 2D images simplifies the process of data analysis and annotation. Many well-established computer vision algorithms and techniques are specifically designed for processing 2D images. Furthermore, LiDAR point clouds can contain millions of points, leading to high-dimensional data. By converting the point clouds to 2D images, the dimensionality is reduced, simplifying the data representation, and potentially reducing computational complexity.

This question may arise that why was LiDAR data initially chosen over 2D images, and why wasn't a 2D dataset used from the beginning? One reason is LiDAR operates based on active sensing, emitting laser beams and measuring the time it takes for them to bounce back. This makes LiDAR less dependent on lighting conditions and immune to issues such as shadows, reflections, or changes in illumination. In contrast, images can be affected by variations in lighting and weather conditions, making them less reliable in certain scenarios. Moreover, in Mobile LiDAR Systems (MLS), LiDAR sensors are mounted on a vehicle's roof. These sensors collect point cloud data as the vehicle is moving. Unlike satellite views,

which provide an overhead perspective, LiDAR data offers a unique viewpoint from ground level. This perspective enables LiDAR to effectively address challenges related to road markings occlusion caused by objects such as, overpasses (Figure 11), and dense vegetation (Figure 12), particularly in regions like Hawaii with lush foliage. This capability enhances the completeness of road marking visualization.
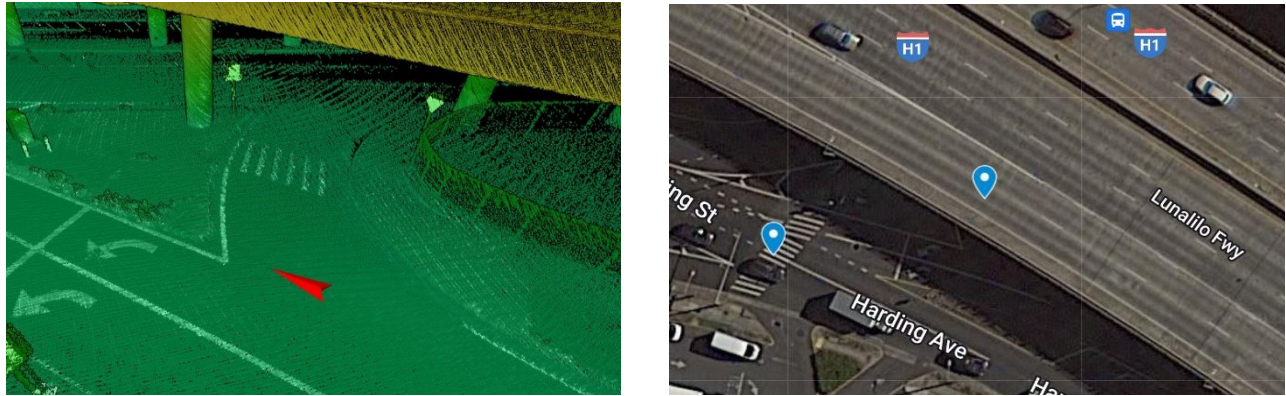
Figure 12. Comparing road marking visibility: (a) LiDAR data vs (b) satellite view with and overpass barrier. An example of a detected crosswalk which is hidden under the overpass from satellite view

(a)   (b)
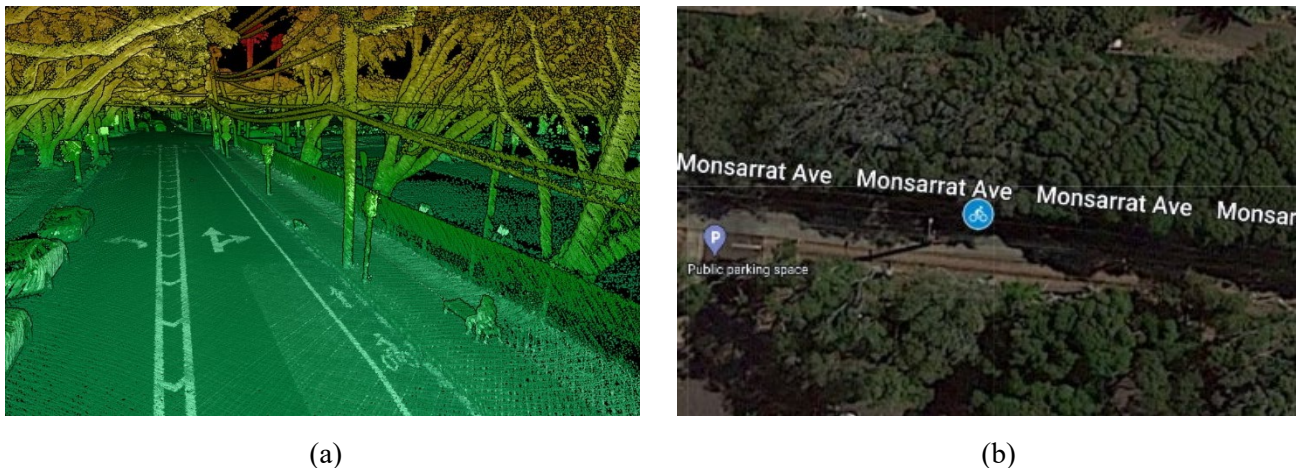
Figure 13. Comparing road marking visibility: (a) LiDAR data vs (b) satellite view with a vegetation barrier. An example of a detected bike sign which is hidden under the trees from satellite view

Step 4 (Annotation): At this stage, our dataset comprises 2D gray-scale images, forming the foundation for the subsequent object annotation step. To annotate the road markings within images, a team deployed to draw 2D polygon bounding boxes around the objects of interest, using the advanced annotation tool called Roboflow, as shown in Figure 13, we enclosed road markings inside axis-aligned bounding boxes. The annotated bounding boxes were then used to generate a labeled dataset, including the raw data of the bounding boxes, which is presented in Table 4.

Table 4. Raw Data of the Bounding Boxes Annotated in Figure 13 provided in Roboflow

RAW DATA		
box	crosswalk	bike sign
label	2	3
center_x	273.6	37.5
center_y	106.25	213.45
width	87.6	53.21
height	191.37	31.49
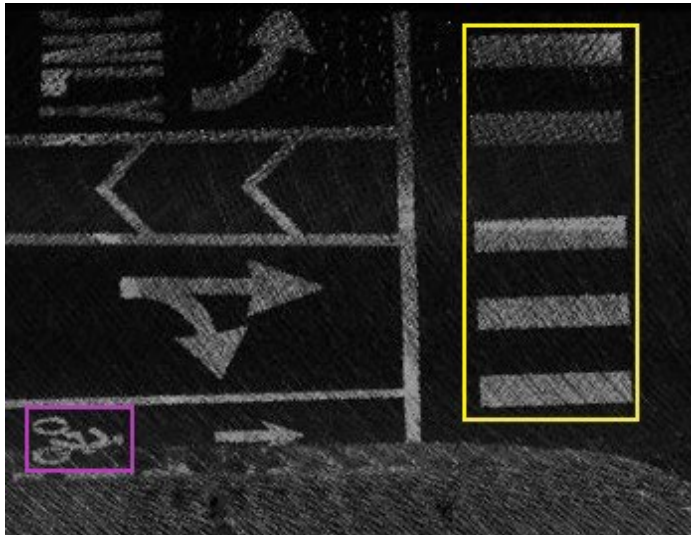
Figure 14. Annotated Intensity Image of a Road Segment in Roboflow

Roboflow, a computer vision platform, provides an intuitive interface for annotating and labeling images. Its annotation capabilities include the ability to create 2D polygon bounding boxes, which are bound the identification of road markings inside images. These bounding boxes are paired with labels, where each label represents a specific road marking class number. Roboflow's annotation management system stores image annotations as structured data in formats like JSON, Pascal VOC XML, or YOLO TXT, capturing bounding box coordinates, class labels, and metadata for each annotated object.

To align with our chosen object detection algorithm, YOLO, we adopted the YOLO TXT format for annotations. In this format, each image is associated with a single text file containing a line for each bounding box annotation. The row format follows: "class_id center_x center_y width height". The fields are separated by spaces, and the coordinates are normalized between zero and one.

This labeled dataset serves as the training data for our object detection model. Due to the time-consuming and labor-intensive nature of object annotation, it poses a significant challenge in object

detection projects. As a result, our current focus is primarily centered around crosswalks and bike signs. However, as we progress, we aim to expand our annotation efforts to include a wider range of road marking types.

Step 5: The next step is dividing the dataset into three subsets for training, evaluation, and testing as follow:

Training Set: 70% (838 images); Evaluation Set: 20% (240 images)

Testing Set: 10% (120 images): The choice of splitting the dataset is based on common practices in machine learning. The larger portion allocated to the training set (70%) allows the model to capture the underlying patterns and characteristics of the dataset. The testing set (10%) serves as an independent benchmark to assess the model's ability to generalize to unseen data. It provides an unbiased evaluation of the model's performance.

Evaluation Set (20%): Serves as a dedicated subset for assessing the model's performance. This subset helps in fine-tuning the model's parameters, optimizing its performance, and making informed decisions about any necessary adjustments during the training process.

3.2.2. Training and Evaluating the Model

This stage involves using YOLO (You Only Look Once), a real-time object detection algorithm that can detect objects in an image and provide bounding boxes and class probabilities for each object. The YOLO algorithm works by dividing an input image into a grid and making predictions for objects within each grid cell. Here is a summary of how YOLO detects objects:

A. Architecture: YOLOv5 consists of a series of convolutional layers, which are responsible for extracting features from the input image. These convolutional layers are followed by up-sampling and down-sampling layers, which help capture features at different scales and resolutions. This multi-scale feature extraction enables the model to detect objects of various sizes in the input image.

B. Grid Division: The input image is divided into a grid of cells. Each cell is responsible for predicting objects that fall within its boundaries. The size of the grid depends on the chosen configuration of YOLOv5 (e.g., YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x). A common configuration, for instance, is a 32x32 grid.

C. Anchor Boxes: Prior to training, anchor boxes are defined. These anchor boxes represent a set of predefined bounding box shapes with different aspect ratios and sizes. Each grid cell predicts bounding boxes relative to these anchor boxes. The anchor boxes help in handling objects of various sizes and aspect ratios.

D. Bounding Box Prediction: For each grid cell, YOLO predicts multiple bounding boxes. Each bounding box consists of coordinates (x, y) representing the box's center, width (w), height (h), and a confidence score indicating the likelihood of an object being present within the box.

E. Object Confidence: The confidence score represents the probability that an object is present in a predicted bounding box. It is computed based on the intersection over union (IoU) between the predicted box and the ground truth box during training. This score is used to filter out low-confidence detections.

F. Class Prediction: Alongside the object confidence scores, YOLOv5 also predicts class probabilities for each bounding box. These probabilities represent the likelihood of each detected object belonging to a specific class, such as a crosswalk of bike sign.

G. Non-Maximum Suppression: To eliminate redundant or overlapping bounding box predictions, YOLO applies a technique called non-maximum suppression. This process removes duplicate detections and retains only the most confident bounding box for each object.

H. Output: The final output of YOLOv5 is a list of bounding boxes, each associated with a class label and a confidence score.

YOLO primarily detects objects within axis-aligned rectangular regions. An axis-aligned bounding box is defined by its top-left and bottom-right corners, and it aligns with the image's coordinate axes. For compatibility with our chosen object detection algorithm, YOLOv5, we also utilize non-rotated vertical oriented rectangle bounding boxes for annotation stage.

We trained the model using the annotated dataset created in the previous stage and evaluated its performance using several metrics, including training loss, validation loss, recall, precision, and mean Average Precision (mAP). The loss function comprises of box loss, objectness, and classification loss.

Table 5. Model Performance Metrics during Training and Validation Stages

Train			Validation			All Road Markings			
box_loss	obj_loss	cls_loss	box_loss	obj_loss	cls_loss	precision	recall	mAP_0.5	mAP_0.5:0.95
0.017342	0.0083676	0.00036619	0.020169	0.0055445	0.0038446	0.92253	0.79529	0.848	0.70291

Before delving into the details, it is worthwhile to refresh our understanding of the meanings of several key terms:

- True positives (TP) (#): The number of objects where the model predicts label
- True Negatives (TN) (#): The number of objects where the model does not predict a label and the ground truth show the object not having a label
- False Positives (FP): The model predicted a label, but it is not a part of the ground truth.
- False Negatives (FN): The model does not predict a label, but it is part of the ground truth.
- mAP: is calculated as the mean of the Average Precision (AP) values across multiple classes. AP is a metric that considers the precision-recall trade-off for a specific class. For instance, when precision increases and recall decreases, it means that the model is becoming more conservative in its predictions. It becomes more selective in detecting positive instances, resulting in fewer false positives but also missing some true positive instances. In this scenario, the precision may

improve due to the reduction in false positives, while the recall decreases due to missing some positive instances.

The following provides a summary of each metric:

1. Train/box_loss: This metric measures the loss associated with bounding box regression during the training phase. Bounding box regression refers to the process of predicting the coordinates (box's center) and sizes (width, and height) of the bounding boxes that tightly enclose objects in an image. The YOLOv5 algorithm aims to minimize this loss to accurately predict the positions of objects in an image.

2. Train/obj_loss: This metric represents the loss associated with object detection during the training phase. It measures how well the algorithm is able to detect and localize objects within an image. Obj_loss is calculated based on the confidence scores assigned to each predicted bounding box. These confidence scores indicate the algorithm's level of certainty regarding the presence of an object within each bounding box.

3. Train/cls_loss: This metric measures the loss associated with object classification during the training phase. It quantifies how accurately the algorithm assigns class labels to the detected objects. Each bounding box prediction is associated with a class probability distribution, indicating the likelihood of the object belonging to different predefined classes (e.g., crosswalk, bike sign, etc.)

4. Precision: Precision (Eq.1) is a commonly used evaluation metric in object detection algorithms. It measures the proportion of correctly predicted positive instances (true positives) out of the total instances predicted as positive (true positives + false positives). A higher precision indicates a lower false positive rate.

$$Precision = \frac{TP}{TP + FP} \ (Eq.1)$$

5. Recall: Recall (Eq.2) is another commonly used evaluation metric in object detection algorithms. It measures the proportion of correctly predicted positive instances (true positives) out of the total positive instances present in the dataset (true positives + false negatives). A higher recall indicates a lower false negative rate.

$$Recall = \frac{TP}{TP + FN} \ (Eq.2)$$

6. mAP_0.5: Mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5. It means that the mAP is computed by evaluating the precision and recall of object detection at an IoU threshold of 0.5. IoU is a measure of the overlap between predicted bounding boxes and ground truth annotations. This threshold determines how much overlap is required for a

predicted bounding box to be considered a true positive. mAP_0.5 indicates how well the algorithm detects and localizes objects at a relatively lenient IoU threshold.

7. mAP_0.5:0.95: Mean Average Precision (mAP) over a range of IoU thresholds from 0.5 to 0.95: This metric provides a more comprehensive evaluation of the algorithm's performance by considering multiple IoU thresholds. Instead of using a single IoU threshold like in the previous metric, mAP_0.5:0.95 evaluates the algorithm's performance across a range of IoU thresholds from 0.5 to 0.95. This range encompasses a wider spectrum of overlap criteria between the predicted and ground truth bounding boxes.

8. val/box_loss: Similar to train/box_loss, this metric represents the loss associated with bounding box regression during the validation or evaluation phase.

9. val/obj_loss: Similar to train/obj_loss, this metric represents the loss associated with object detection during the validation or evaluation phase.

10. val/cls_loss: Similar to train/cls_loss, this metric represents the loss associated with object classification during the validation or evaluation phase.

Monitoring and analyzing these metrics help us understand how well the model localizes objects, detects them accurately, assigns correct class labels, and balances precision and recall. By optimizing these metrics, we can improve the performance of the model and make it more reliable and effective in object detection tasks.

3.2.3. Model Results

The model achieved the mAP_0.5 score of 0.848 which suggests that the model was able to accurately detect almost 85% of the objects in the training dataset that had an IoU of 0.5 or higher for detecting both crosswalk and bike signs. Additionally, the results for crosswalks and bike signs separately show that the model is particularly strong in detecting crosswalks, achieving a mAP_0.5 score of 0.924. The mAP_0.5 score for bike signs is also relatively high, at 0.773.

The mAP_0.5:0.95 score of 0.709 indicates that the model performed slightly less accurately for objects with a wider range of IoU values (ranging from 0.5 to 0.95), but still achieved a relatively high level of accuracy. The mAP_0.5:0.95 score of 0.821 for crosswalks and 0.604 for bike signs achieved. It is generally observed that object detection models, including YOLO models, tend to demonstrate higher performance at lower IoU thresholds since as the IoU threshold is increased, the model requires a higher degree of overlap with the ground truth annotations to classify a detection as a true positive.

The higher mAP value for crosswalks compared to bike signs can be attributed to the larger dataset size of crosswalks (800 images) compared to bike signs (400 images), allowing model to learn crosswalks specific patterns and characteristics more effectively. Additionally, the relatively simpler shape of crosswalks compared to bike signs likely contributed to the model's higher accuracy in detecting crosswalks.

After reviewing the training and validation results, we now shift our focus to the testing results presented in Table 6, illustrating the metrics of precision, recall, mean Average Precision (mAP), and F1 (the measure of the harmonic mean of precision and recall) score. These metrics were evaluated using an IoU threshold of 0.5, which determines the level of overlap between predicted bounding boxes and the ground truth. If the IoU value is above a certain threshold, the predicted bounding box is considered a true positive detection, otherwise, it is classified as a false positive or a false negative. This threshold is often considered adequate for evaluating the accuracy of object localization tasks. The YOLOv5 model utilized in the testing phase is loaded with the pre-trained weights that were obtained during training the model on a separate training dataset. These weights correspond to the best performance of the model based on evaluation criteria such as loss on validation set. In the testing phase, the model performs object detection on each image within the testing dataset, which constitutes 10% of the overall dataset and includes 120 images. For every image processed, the model predicts bounding boxes and assigns confidence scores to different object classes. These confidence scores reflect the model's belief or probability that the corresponding bounding box contains an object of a specific class. To clarify, the IoU threshold is used to determine true positives and calculate precision and recall, while the confidence threshold is used to classify detections as positive or negative based on their confidence scores. Additionally, The F1 score is calculated as the harmonic mean of precision and recall (Eq.3). It provides a balanced measure of the model's accuracy, incorporating both precision and recall into a single value.

$$F_1 = 2 * \frac{Precison * Recall}{Precision + Recall} \; (Eq. 3)$$

Table 6. Testing Results

ROAD MARKING TYPE	TP	FP	FN	PRECISION	RECALL	F1 SCORE	MAP_0.5
Crosswalk	79%	3%	18%	0.963	0.814	0.882	0.889
Bike Sign	71%	0%	29%	1	0.71	0.83	

Figure 14, along with Table 6, serves as a comprehensive representation of the model's performance across 100 epochs during the training and evaluation process. It provides a visual depiction of the model's progression. During the initial epochs, the model's performance may be relatively poor, with high values for box_loss, obj_loss, and cls_loss. However, as training progresses, these loss values gradually decrease, indicating improved accuracy in object localization, objectness prediction, and object classification. Over time, precision and recall values increase, resulting in fewer false positives and false negatives. The mAP_0.5 and mAP_0.5:0.95 values also improve as the model's performance advances across IoU thresholds and object classes.
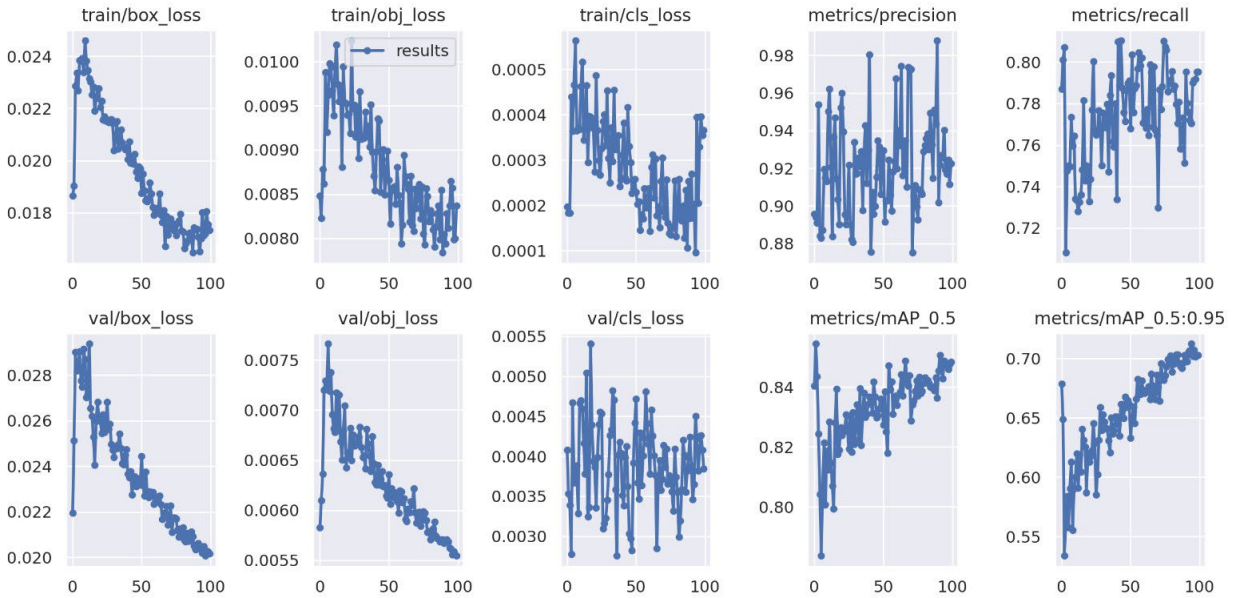
23

Figure 15. Model Performance Progression across Training and Evaluation through 100 epochs

Figures 15, 16, and 17 showcase Performance Analysis across Confidence Thresholds. In F1-confidence Curve, the F1 score for all classes reaches a value of 0.85 at a confidence threshold of 0.684, this point on the curve represents a desirable operating point where both precision and recall are relatively high. In a precision-confidence curve, the precision for all classes reaches a value of 1 at a confidence threshold of 0.908, indicating that all the detections made by the model at this threshold are correct. In a recall-confidence curve, recall for all classes is 0.88 when the confidence threshold is set to 0.00. Recall measures the ability of the model to correctly detect positive instances. This suggests that the model is capturing a relatively high proportion of positive instances, but it may also result in a higher number of false positives.
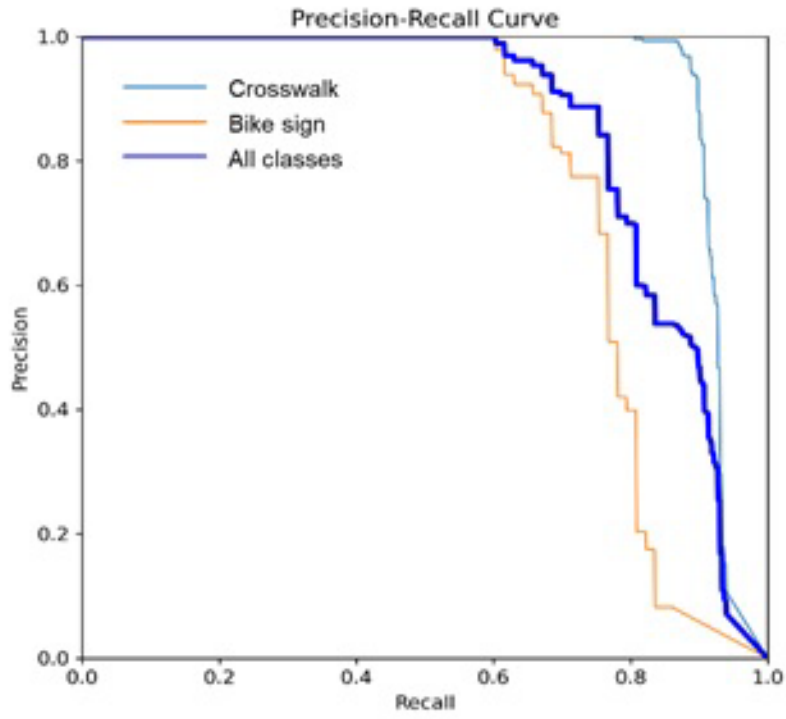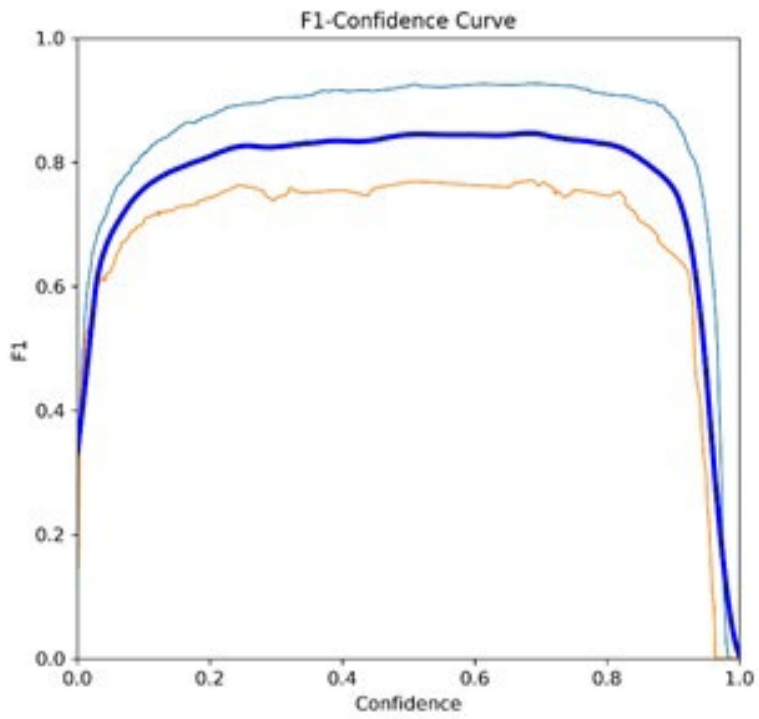
Figure 16. Precision-Recall Curve
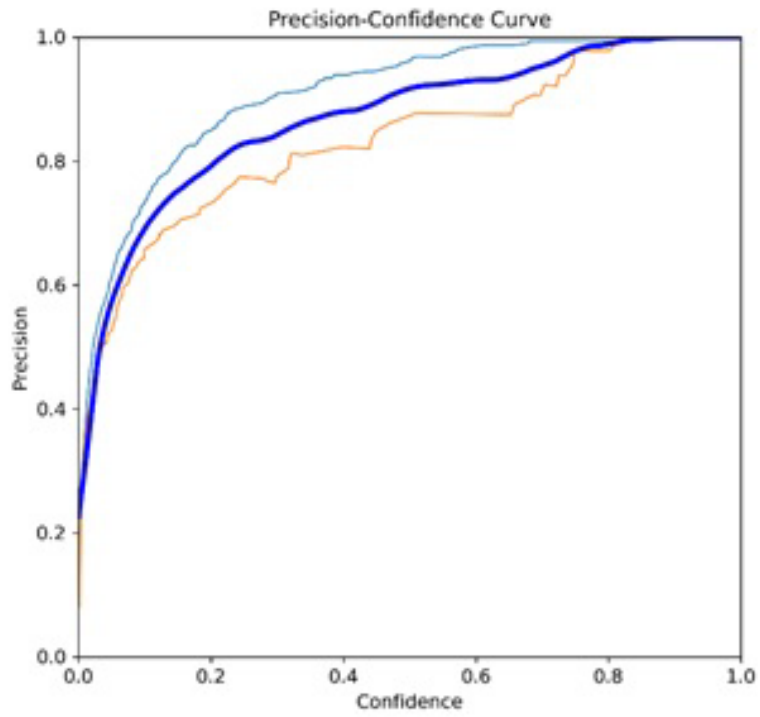

Figure 17. F1-Confidence Curve

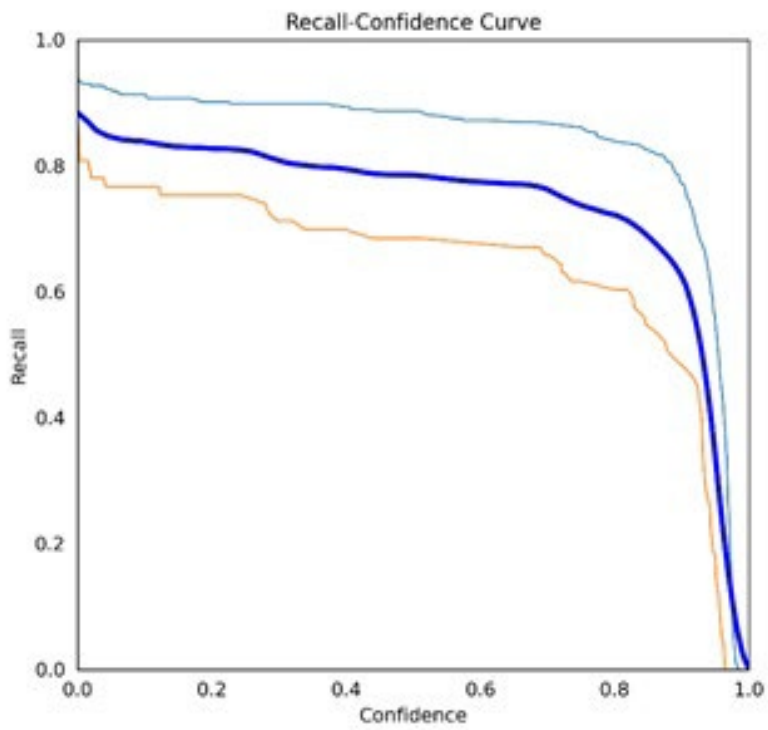Figure 18. Precision-Confidence Curve
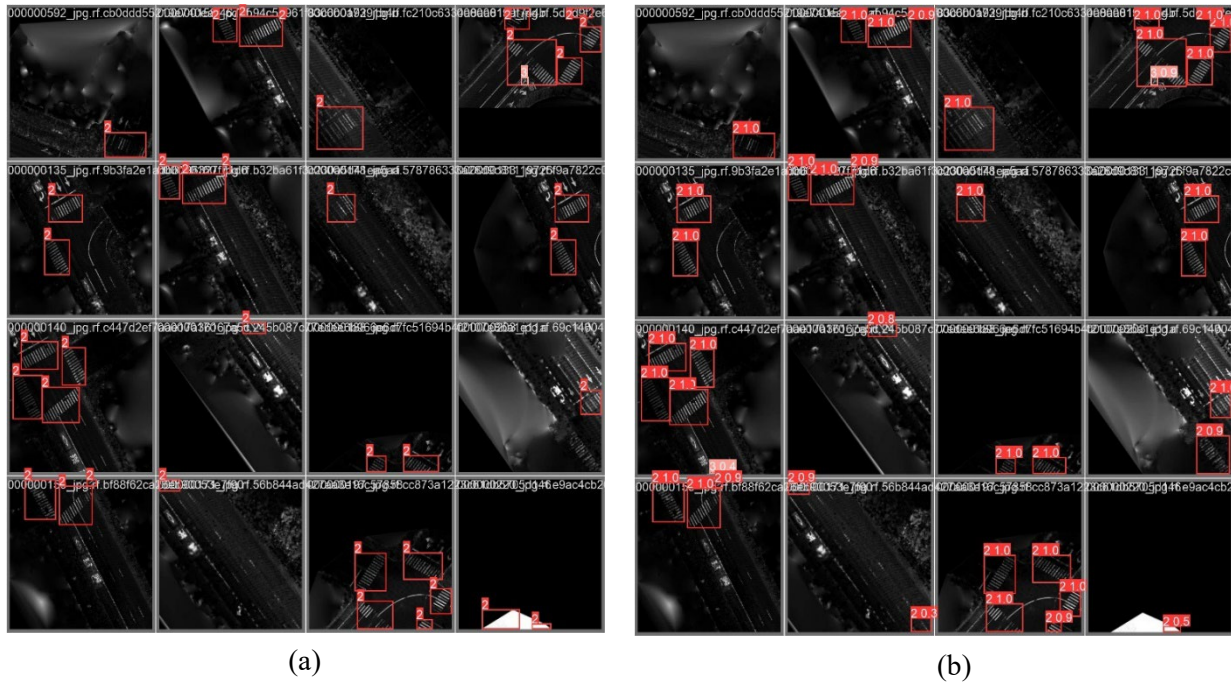

Figure 19. Recall-Confidence Curve

Figure 20. Sample Representation of Object Detection with YOLO Algorithm – a) Annotated bounding box with ground truth labels b) corresponding predicted bounding boxes, labels, and confidence scores.

3.2.4. Geo-Locating

In order to accurately locate road markings in their correct spatial position on the Earth's surface, a transformation process is employed to convert the pixel coordinates of an image to real-world coordinates. This transformation relies on the geo-referenced information contained in an auxiliary XML file generated during the conversion of LiDAR data to a raster image. The auxiliary XML file includes geo-referencing parameters such as the pixel size, rotation angle, and the coordinates of the upper-left corner of the image. These parameters relate the pixel coordinates of a raster image to real-world geographic coordinates, and provide the starting point for our conversion process. First, we need to determine the pixel coordinate of the road marking within the image (The top-left pixel of the image is often assigned coordinates (0, 0)), we employ the pre-trained YOLOv5 model to predict the geometric properties of object bounding boxes. These properties include the top-left and bottom-right pixel coordinates of each bounding box. By leveraging this information, we can accurately identify the pixel location of the midpoint of the road marking. Once we have determined the pixel coordinate of the road marking within the image we can utilize the projected coordinates of the top-left corner of the image, along with the pixel size and rotation parameters, to project the road markings' coordinates onto the raster files' projected coordinates. Finally, we used a defined projection system with UTM zone 4 and the WGS84 ellipsoid to convert projected coordinates of the road markings to latitude and longitude, representing the geographic location on the Earth's surface.

The following is an explanation of the projection system:

27

- UTM: Universal Transverse Mercator is a commonly used map projection system that divides the Earth into zones, each with its own projection parameters. UTM projections provide accurate measurements of distances and angles within each zone.

- Zone: This parameter specifies the UTM zone number, which represents the specific zone within which the geographic coordinates fall. The UTM projection divides the Earth into 60 zones, each covering a strip of longitude. In this case, the choice of zone=4 indicates that the UTM zone being used is zone 4. This selection aligns with the dataset's geographic coverage of Hawaii roads and highways since Hawaii falls within UTM zone 4.

- Ellipsoid: This is the ellipsoid used for the projection. The Earth is not a perfect sphere, so map projections typically use an ellipsoidal model to approximate its shape. WGS84 (World Geodetic System 1984) is a widely used ellipsoid model commonly associated with GPS and global mapping.

These parameters together define the projection system used for the transformation. In this case, the UTM projection is employed, specifically for UTM zone 4, with coordinates referenced to the WGS84 ellipsoid. The output of this geo-locating section is a comprehensive map of crosswalks (31) and bike signs (32), which will be discussed in more detail. The map provides a comprehensive layer of crosswalk and bike sign locations within streets for which LiDAR data has been generated. The aerial overview of crosswalk and bike sign maps are illustrated in Figure 20 and Figure 21 respectively. It worth mentioning that the map is a work in progress and not yet complete. We are committed to continuously updating and expanding the map to include all existing road markings within the city boundaries.
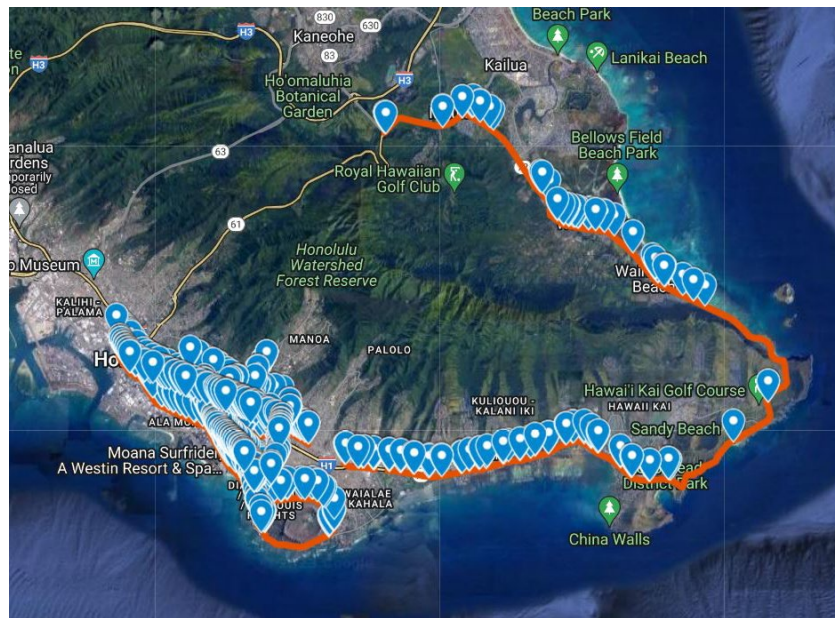

Figure 21. Aerial Overview of Crosswalks Map Including Source Street Boundaries
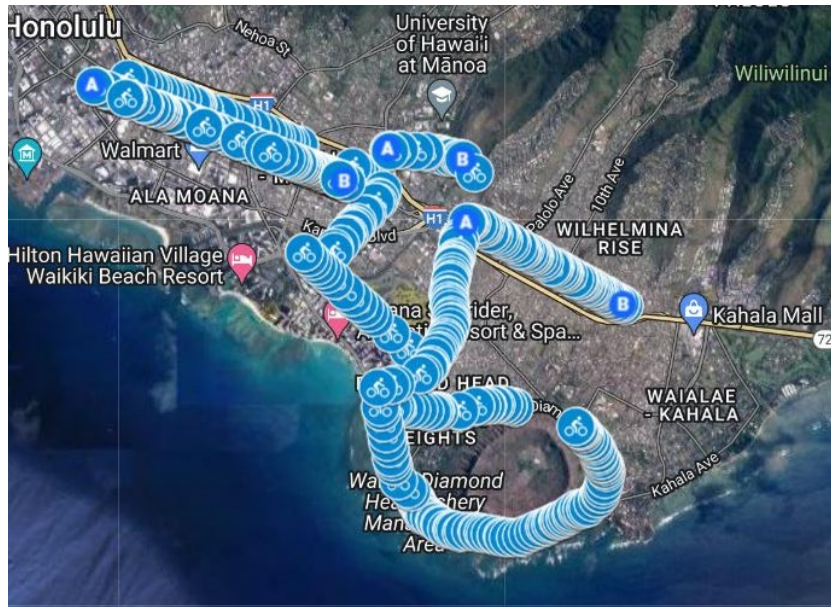
Figure 22. Aerial Overview of Bike Signs Map Including Bike Lanes

The City Map stands out from other maps in two significant ways:

1. Distinct Representation: Unlike traditional maps where road markings, including crosswalks, are integrated into satellite views or other imagery, in this map, road markings are distinct data points, independent of satellite imagery, and separated from other map elements. In addition, satellite views can be hindered by various visual obstructions, such as vegetation, which may obstruct the visibility of markings on the road surface. By treating road markings as distinct data points, we eliminate these potential barriers.

2. Enhanced Data Integrity: In contrast to other cooperative maps, the City Map minimizes the risk of manual errors associated with manually integrating road markings data into existing maps. Many interactive maps suffer from limited data availability and completeness, requiring significant manual effort to gather and update the information. The City Map overcomes these challenges by directly capturing and presenting crosswalk data as separate and reliable data points. This approach ensures the accuracy and completeness of the road markings information.

3. Enhancing Cycling Infrastructure: The Bike Sign Map, offers an extensive view of bike lanes within streets where LiDAR data has been provided from. After detecting the locations of bike signs, we further determined the routes for bike lanes by connecting the bike signs together. Our ultimate goal is to bridge the gaps in existing bike lane infrastructure and create a more cyclist-friendly network.

In the geo-locating stage, we aimed to quantify the discrepancy between the predicted location of road markings and the ground truth annotations. The analysis revealed an average positional discrepancy of 5.61 feet between the centers of annotated bounding boxes and predicted bounding boxes for

crosswalks. For bike signs, the discrepancy was measured at 2.38 feet. Despite achieving a higher mAP_0.5 score for crosswalks, indicating better IoU performance, compared to bike signs, we observed a larger discrepancy in the geo-location of crosswalks. One potential explanation is the difference in size and shape between the two types of road markings. Crosswalks are typically larger and more irregular in shape, requiring larger adjustments for accurate geo-location. The relatively low values of discrepancy achieved are a testament to the precision and accuracy of our system in determining the spatial locations of road markings.

CHAPTER 4.   FINDINGS AND DISCUSSION

The goal of this project has been to investigate how the application of LiDAR technology for infrastructure asset management and inventories has expanded and become increasingly feasible, due partly to ubiquitous consumer access to LiDAR point clouds, and their emerging use in autonomous vehicle sensor systems. Additionally, many State DOTs have initiated regular collection of point clouds as part of their asset management efforts. In Mobile LiDAR Systems (MLS), LiDAR sensors are mounted on a vehicle's roof. These sensors collect point cloud data as the vehicle is moving.

As an example, the Hawaii Department of Transportation (HDOT) surveys, manages, maintains and inspects around 5,000 miles of highways across six islands in the State. In 2003, HDOT started collecting photo-log images via cameras mounted onboard a mobile data collection vehicle. Starting in 2009, a combined LiDAR/digital camera system was used to measure and record around 1,000 miles of roadway data on Oahu. The resulting point cloud data consisted of roughly 200 billion points (or 630 gigabytes of data). Similar survey efforts were initiated in 2011 and 2012 for the Island of Hawaii and Islands of Kauai and Maui respectively through a contract with Mandli Communications.

Figure 22 below provides an overview of roadways in Hawaii were LiDAR point cloud data has been collected since 2009.
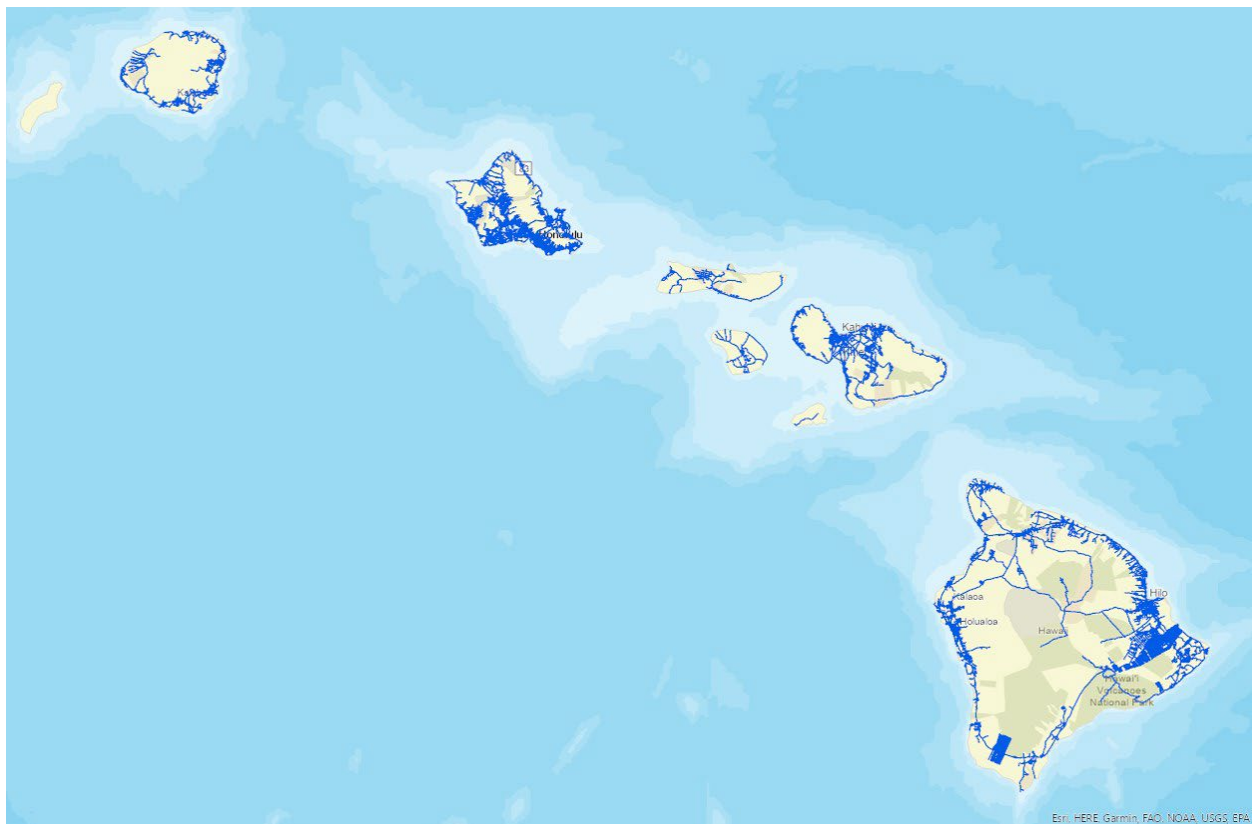

Figure 23. Collected LiDAR point cloud data across State of Hawaii

An example of a high-resolution photo-log image and high-resolution LiDAR point cloud data for the same road intersection (Dole Street and East-West Road on the UH Manoa Campus) are shown in Figures 23 and 24.

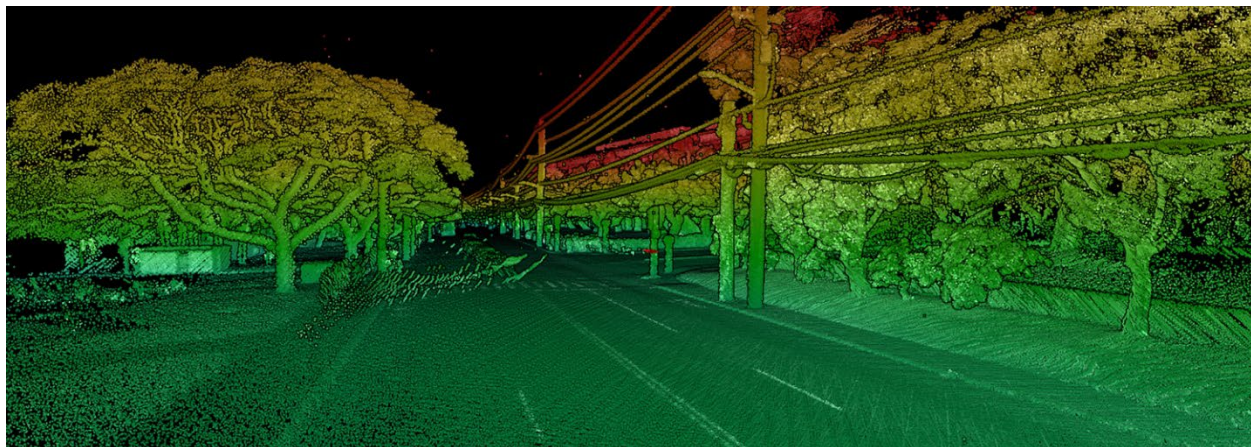Figure 24. High-resolution photo-log image sample

Figure 25. High-resolution LiDAR point cloud data sample

The methods presented in Chapter 3 of this report have shown that the automatic detection and geocoding of relevant roadway assets may be possible through the analysis of LiDAR point cloud data. This approach has two major benefits. First, it eliminates the need for inspections and related work zone safety issues and second, this approach permits quicker detection and response to changing road conditions. Furthermore, attributing geocoded assets with other variables broadens the analysis possible and questions addressed including:

- Safety Analysis: geocoded traffic incident data,

- Census Tract Data: household socio-demographics, community data, and

- Infrastructure Quality: pavement and roadway condition ratings.

As stated above, the main goal of this project has been the detection and geolocation of road markings such as crosswalks and bike symbols on roadways. Furthermore, we investigated the relationship of the presence of these road markings with observed crash rates. The map below shows the overall conceptual framework for the project for the same intersection shown in Figures 23 and 24.
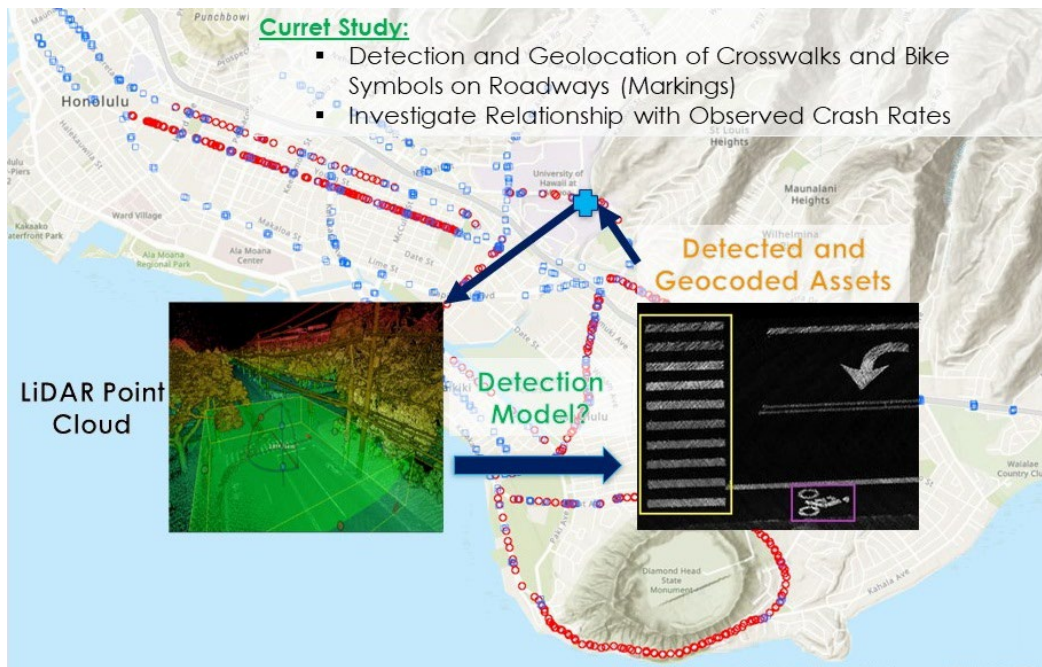

Figure 26. Conceptual framework for road marking detection using LiDAR data

Utilizing LiDAR point cloud data for detection of road markings has certain advantages compared to satellite imagery. Examples of these advantages include detection of markings that would be blocked by other road facilities and natural obstructions such as trees. Figure 26 provides an example of a crosswalk on Kapiolani Boulevard that is blocked from satellite imagery by the H1 Interstate, and Figure 27 is an example of trees blocking a bike symbol. In both instances, the road markings can be detected using LiDAR point cloud data.

(a)   (b)

Figure 27. Crosswalk (a) detected by LiDAR, and (b) obstructed from satellite imagery

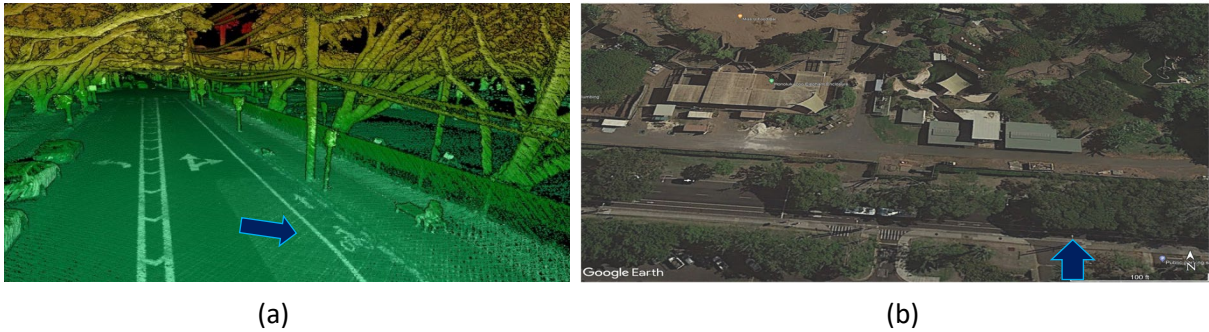(a)   (b)

Figure 28. Bike symbol (a) detected by LiDAR, and (b) obstructed from satellite imagery

4.1. Analysis Framework

The analysis framework developed for this project involved (a) preprocessing of data, through which 2D images were extracted from LiDAR point clouds; (b) developing a detection model using YOLOv5, including training and testing/validation of the model; (c) applying the developed detection model to HDOT point clouds to detect road markings; and (d) using the geocoded locations of roadway symbols for a safety analysis, through a Poisson Regression Analysis on bike and pedestrian incident data. Figure 28 below visualizes the analysis framework.
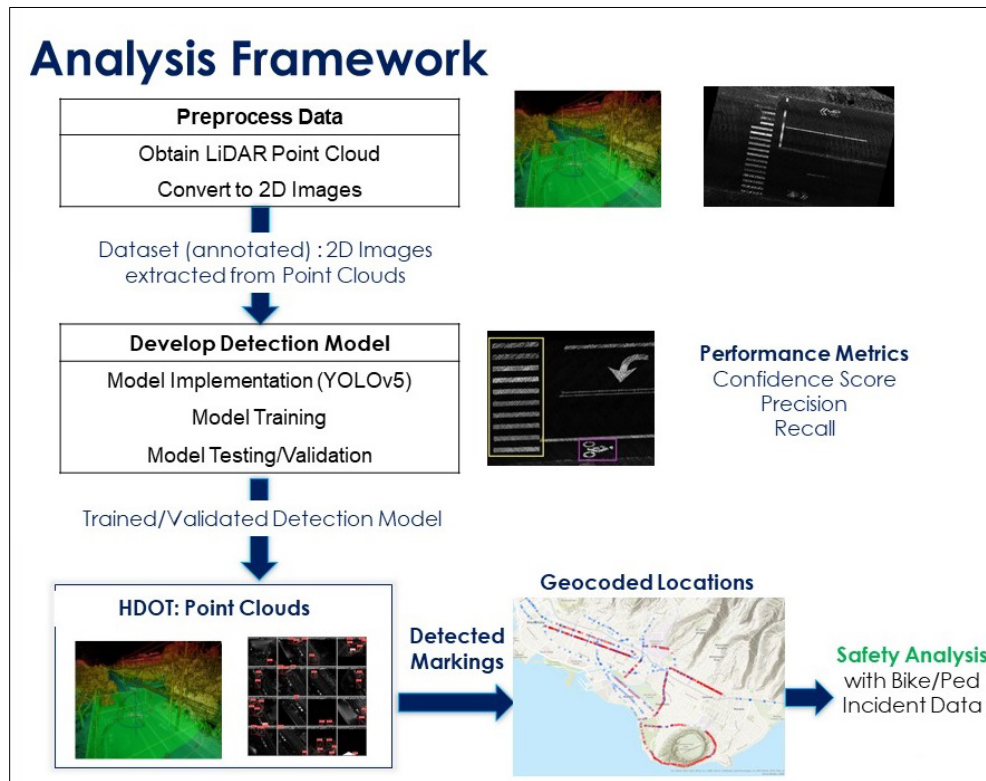

Figure 29. Analysis Framework

4.2. Detection Model

The detection model was YOLOv5 (You Only Look Once) model, an off-the-shelf model implemented in Python based on Neural Network modelling. This model is used to identify objects within a 2D image. A total of 1,198 annotated images were used for model training and testing, of which 70% (838 images) were used for training, 20% (240 images) used for evaluation, and 10% (120 images) for testing. The roadway coverage for the development of the model, including training and testing/validation, covered key streets in the City and County of Honolulu, based on a discussion with Hawaii Department of Transportation (HDOT) and the Department of Transportation Services. Evaluation metrics for this model include:

- Precision = True Positive/(True Positive + False Positives),
- Recall = True Positive/(True Positive + False Negatives), and
- Confidence Score (0-1): Likelihood the object is in the detected box.

A Precision-Recall Curve, showing the Mean Average Precision (area under the P-R curve) was calculated to assess the performance of the model as shown in Figure 29. The results show a detection of nearly 85% objects. However, the model performed better in the detection of crosswalks than bike markings.
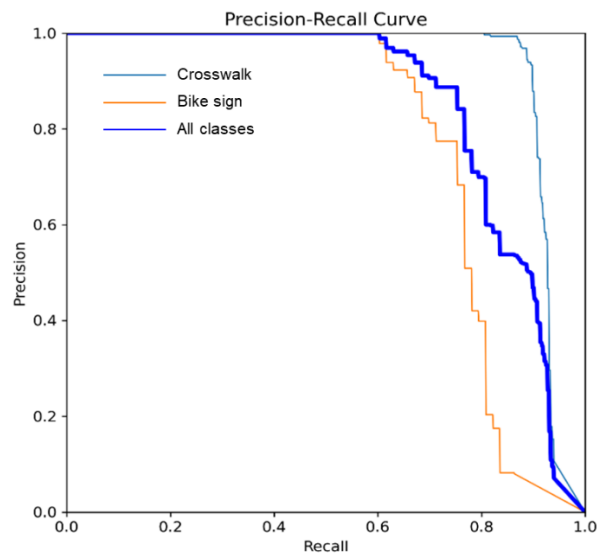

Figure 30. YOLOv5 Model Precision-Recall Curve for 1,198 images

The F1-Confidence Curve shown in Figure 30 signifies that the optimal point for the model is reached at an F1 value of 0.85, corresponding with a confidence threshold of 0.684.

35
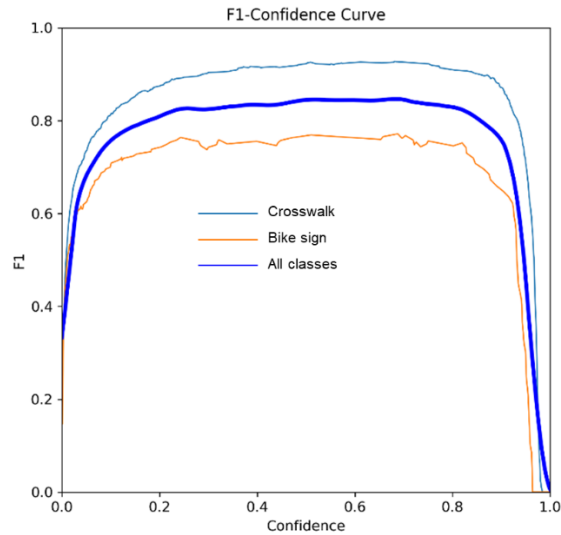
Figure 31. YOLOv5 F1-Confidence Curve for 1,198 images

The next step in the project was to use the detected roadway markings for a safety analysis. Additional roadways from northern and western shores of the island of Oahu, as well the cities of Wahiawa and Haleiwa were included within the analysis framework described above. A total of 1,071 crosswalks and 517 bike symbols were detected throughout this process, as shown in Figures 31 and 32.

Figure 32. Detected crosswalks based on the YOLOv5 model (n = 1,071)

Figure 33. Detected bike symbols based on the YOLOv5 model (n = 517)

4.3. Safety Analysis

Several data sources were used for safety analysis, which included:

- HDOT Point Cloud Data (accessed January 2022)
- Pedestrian-involved and bicycle-involved major traffic crashes (between January 2019 and December 2022) as shown in Figure 33.
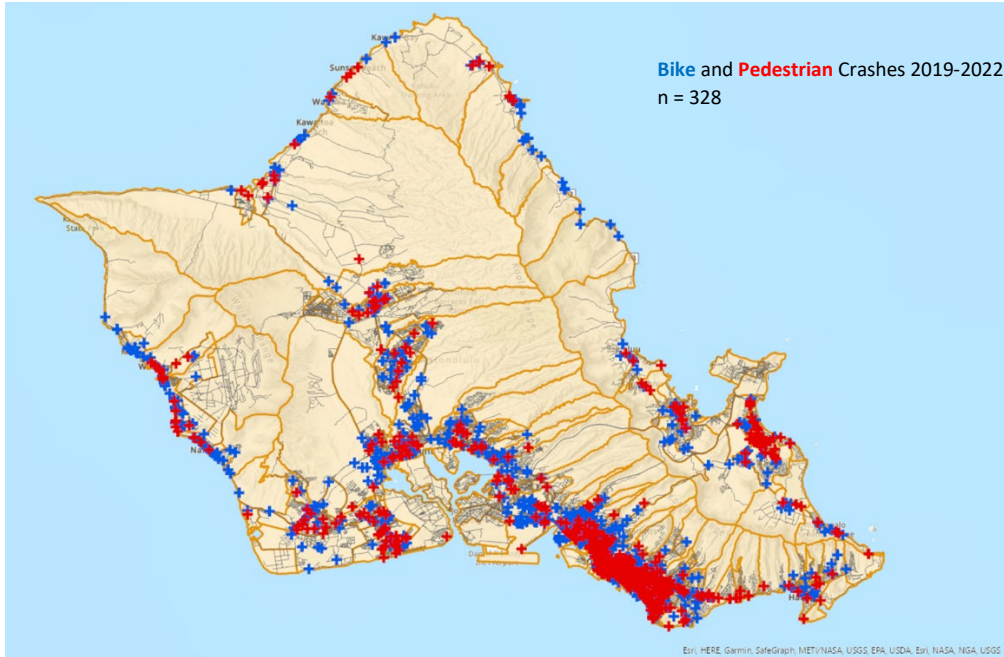- US Census Tract Level Demographics (2022)

Figure 34. Bike (blue) and pedestrian (red) crashes on Oahu 2019-2022 (n = 328)

We used a Poisson Regression Analysis on crash counts to analyze safety across census tracts, with the observed outcomes of number of crashes per Census Tract. Explanatory variables included:

- Number of Identified Assets from the Model
- Model Performance
- Census Tract Demographics

Model estimation results for pedestrian crashes across Census Tracts on Oahu are presented in Table 7:

Table 7. Model estimation results for pedestrian crashes

Variable	Estimate	Std. Error	t-statistic
Constant	2.618	0.5865	4.464
Number of Detected Crosswalks	0.02577	0.003097	8.321
Model Performance: Mean Confidence	-0.1877	0.07153	-2.624
Population density (ppl/km-sq)	-2.6E-05	7.08E-06	-3.611
Total Road Length (km)/Tract Area (km-sq)	0.06028	0.007638	7.893
Hawaiian Homelands (km-sq)/Tract Area (km-sq)	-0.5372	0.3716	-1.446
Number of Census Tracts	108		
Residual Deviance	89.585		

The model shows that Census tracts with a greater detection of crosswalks showed higher crash rates, and Tracts with a higher mean confidence scores (related to marking quality) showed lower crash rates.

Model estimation results for bike crashes across Census tracts on Oahu are presented in Table 8:

Table 8. Model estimation results for bike crashes

Variable	Estimate	Std. Error	t-statistic
Constant	4.424	0.5881	7.523
Number of Detected Bike Symbols	0.01107	0.002705	4.093
Model Performance: Mean Confidence	-0.4445	0.06946	-6.399
Population density (ppl/km-sq)	-7.2E-05	2.57E-05	-2.814
Total Road Length (km)/Tract Area (km-sq)	0.01822	0.01847	0.987
Number of Census Tracts	108		
Residual Deviance	420.7		

Similar to results for pedestrian crashes, the model shows that tracts with a greater detection of bike symbols showed high crash rates, and tracts with detected symbols that have a higher mean confidence (greater marking quality) showed lower crash rates.

Furthermore, it can be concluded that higher population density tracts are associated with lower crash rates, and higher roadway length to tract area ratios are associated with higher crash rates, though this was statistically insignificant for bike crashes. For pedestrian crashes, tract areas with higher areas of Hawaiian Homeland per tract area show lower pedestrian crash rates, but this too was not statistically significant.

CHAPTER 5.   CONCLUSIONS

LiDAR is a technology for 3D mapping that is widely used in civil infrastructure asset management and inventories. Several State DOTs, including Hawaii State Department of Transportation (HDOT), have taken the advantage this to collect point clouds for transportation infrastructure asset management. This study demonstrates an application of LiDAR point clouds in identifying non-motorized infrastructure elements, focused on developing an automated method for extracting, detecting, and geo-locating road markings from LiDAR point clouds. The utilization of LiDAR technology provided several advantages, including robustness to lighting conditions, and overcoming visual barriers.

ArcGIS Pro was used for data preprocessing, which including ground identification, intensity image generation, and training sample preparation. A YOLOv5 object detection model was trained for road marking detection on the rasterized point clouds. Through the training of the YOLOv5 object detection algorithm, the model achieved promising results in accurately detecting and localizing road markings, including crosswalks and bike signs. The mAP_0.5 score of 0.848 showcases an overall accuracy of approximately 85% for detecting road markings with an IoU threshold of 0.5 or higher.

The geo-transforming process enabled precise spatial location of road markings on the Earth's surface, the average positional discrepancy error between the centers of predicted and ground truth bounding boxes was measured to be 5.61 feet for crosswalks and 2.38 for bike signs. There are limitations in this study. The attributes available for analysis in LiDAR dataset was limited, the manual labeling process would also consume a high number of man-hours. The long-term objective of study is to develop an automatic infrastructure mapping and surveying process, which will be convenient for engineering applications. Future studies will consider a more robust and widely applicable LiDAR based infrastructure detection system. A more efficient data labeling process and infrastructure LiDAR database also need to be built.

Overall, these results demonstrate our success in minimizing the discrepancy between predicted and ground truth locations, indicating the effectiveness of our geo-locating approach. The generated comprehensive map of crosswalks and bike signs provides a distinct representation of road markings which eliminates visual obstructions and minimizes the risk of errors associated with manually integrating road markings into existing maps.

Over time, the availability of data for digital models of infrastructure systems that support active travel has improved. This includes an increasing availability of LiDAR point clouds, high-resolution photogrammetry, crowdsourcing of GPS data, and geocoded community feedback (such as tweets, smartphone photos, etc. However, several barriers and issues remain before these models and related data can be fully harnessed. The necessary workflow for model creation and analysis is complicated. For example, processing of geospatial data can be a daunting task. Furthermore, objects differ in difficulty of detection due to complexity in geometry, such as crosswalks and bike symbols. In addition, automated detection using deep learning requires large volumes of data training. Finally, as with similar studies, issues related to representativeness of models still exist.

There are limitations to this study. As the attributes available for analysis in the LiDAR dataset was limited, the manual labeling process also consumed a large number of man-hours. The long-term objective of the study was to develop an automatic infrastructure mapping and surveying process, which will be convenient for engineering applications. Future studies will consider a more robust and widely

applicable LiDAR based infrastructure detection system. A more efficient data labeling process and infrastructure LiDAR database also need to be built.

CHAPTER 6.   REFERENCES

1. Jung, J., Olsen, M. J., Che, E., Parrish, C., & Pacific Northwest Transportation Consortium. (2020). Efficient Extraction and Evaluation of Complex Pavement Markings from Mobile Laser Scan Data.

2. A plan for every section of every road on every island. https://onlinepubs.trb.org/onlinepubs/Conferences/2018/AssetManagement/SulijoadikusumoAPlanfor EverySectionofEveryRoad.pdf

3. Morton, J. R. M., Uchida, D. U., Babcock, R. B., & Kozlov, A. K. (2021). Honolulu Complete Streets Annual Progress Report for Fiscal Year 2021. City and County of Honolulu. https://www.honolulu.gov/rep/site/dts/Honolulu_Complete_Streets_Progress_Report_FY2021.pdf

4. About Mandli Communication. https://www.mandli.com/

5. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

6. Hoffmann, G.M.; Tomlin, C.J.; Montemerlo, M.; Thrun, S. (2007). Autonomous automobile trajectory tracking for off-road driving: Controller design, experimental validation, and racing. In Proceedings of the American Control Conference, New York, NY, USA, 9–13 July 2007 (https://ieeexplore.ieee.org/abstract/document/4282788)

7. Holgado-Barco, A.; Riveiro, B.; González-Aguilera, D.; Arias, P. (2017). Automatic inventory of road cross-sections from mobile laser scanning system. Comput. Aided Civ. Infrastruct. Eng., 32, 3–17. (https://onlinelibrary.wiley.com/doi/10.1111/mice.12213)

8. Loukaitou-Sideris, Anastasia; Liggett, Robin; Sung, Hyun-Gun (2007). Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles, Journal of Planning Education and Research 26:338-351. (https://escholarship.org/content/qt7pc652ws/qt7pc652ws.pdf)

9. D.A. Pomerleau, Alvinn (1989). An autonomous land vehicle in a neural network, in: Advances in Neural Information Processing Systems. (https://proceedings.neurips.cc/paper/1988/file/812b4ba287f5ee0bc9d43bbf5bbe87fb-Paper.pdf)

10. Kitawaki, Suguru; Wu, Haiyuan; Shioyama, Tadayoshi (2002). Length Estimation of Crosswalk from Image Data (https://www.jstage.jst.go.jp/article/sicetr1965/38/5/38_5_495/_article/-char/en/)

11. Se, Stephen; Brady, Michael (2003). Road feature detection and estimation in: Machine Vision and Applications (http://stephense.com/research/papers/mva03.pdf)

12. LeCun Y, Bengio Y, Hinton G (2015). Deep learning. Nature 521(7553):436–444 (https://doi.org/10.1038/nature14539)

13. Z.S. Chen, D.F. Zhang (2017). An effective detection algorithm of zebra-crossing, in: International Conference on Electrical and Information Technologies for Rail Transportation, Springer. (https://link.springer.com/chapter/10.1007/978-981-10-7986-3_81)

14. K. Ashraf, et al. (2016). Shallow networks for high-accuracy road object-detection. (https://arxiv.org/abs/1606.01561)

15. W. Ouyang, X. Wang (2013). Joint deep learning for pedestrian detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2013 (https://openaccess.thecvf.com/content_iccv_2013/papers/Ouyang_Joint_Deep_Learning_2013_ICCV_paper.pdf)

16. Malbog, Mon Arjay (2019). MASK R-CNN for Pedestrian Crosswalk Detection and Instance Segmentation, in 6th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), 2019 (https://ieeexplore.ieee.org/abstract/document/9117217)

17. Redmon J, et al. (2016). You only look once: unified, real-time object detection. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR). (https://doi.org/10.1109/cvpr.2016.91)

18. Bochkovskiy A, Wang CY, Liao HYM (2020). YOLOv4: optimal speed and accuracy of object detection, p 1. (https://arxiv.org/abs/2004.10934v1)

19. Ultralytics (2020). YOLOv5. 2021–02–01). https://github.com/ ultralytics/yolov5/tree/v4.0.

20. Zhang, Zheng-De; Tan, Meng-Lu; Lan, Zhi-Cai, Liu, Hai-Chun; Pei, Ling; Yu, Wen-Xian (2022). CDNet: a real-time and robust crosswalk detection network on Jetson nano based on YOLOv5, in: Neural Computing and Applications (https://link.springer.com/article/10.1007/s00521-022-07007-9)

21. K. Rebai, N. Achour, O. Azouaoui (2014). Road intersection detection and classification using hierarchical SVM classifier, Adv. Robot. 28 (14), 929–941. (http://refhub.elsevier.com/S0378-4371(19)31958-2/sb7)

22. Hata, Alberto; Wolf, Denis (2014). Road Marking Detection Using LIDAR Reflective Intensity Data and its Application to Vehicle Localization; in: 2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC), 2014 (https://ieeexplore.ieee.org/abstract/document/6957753)

23. B. Riveiro, et al. (2015). Automatic detection of zebra crossings from mobile LiDAR data, Opt. Laser Technol. 70, 63–70 (https://www.sciencedirect.com/science/article/abs/pii/S0030399215000237)

24. Y. Zhang, et al. (2015). 3D liDAR-based intersection recognition and road boundary detection method for unmanned ground vehicle, in: Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, IEEE, 2015. (https://ieeexplore.ieee.org/abstract/document/7313180)

25. Q. Zhu, et al. (2012). 3D LiDAR point cloud based intersection recognition for autonomous driving, in: Intelligent Vehicles Symposium (IV), 2012 IEEE, IEEE, 2012. (https://ieeexplore.ieee.org/abstract/document/6232219)

26. Soilán, M., Riveiro, B., Martínez-Sánchez, J., & Arias, P. (2017). Segmentation and classification of road markings using MLS data. ISPRS Journal of Photogrammetry and Remote Sensing, 123, 94-103.

27. Yao, L., Chen, Q., Qin, C., Wu, H., & Zhang, S. (2018). Automatic extraction of road markings from mobile laser-point cloud using intensity data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 42, 3.

28. Chen, S., Zhang, Z., Zhong, R., Zhang, L., Ma, H., & Liu, L. (2020). A dense feature pyramid network-based deep learning model for road marking instance segmentation using MLS point clouds. IEEE Transactions on Geoscience and Remote Sensing, 59(1), 784-800.

29. Wen, C., Sun, X., Li, J., Wang, C., Guo, Y., & Habib, A. (2019). A deep learning framework for road marking extraction, classification and completion from mobile laser scanning point clouds. ISPRS journal of photogrammetry and remote sensing, 147, 178-192.

30. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

31. City Crosswalks Map.
(https://www.google.com/maps/d/edit?mid=1xq_JrJiS_mEkx-YKI09mhEOemKjCTYU&usp=sharing)

32. City Bike Signs Map.
(https://www.google.com/maps/d/edit?mid=1kYd5ICkJ3BsGSoi5q8Y9TfR3PkaNOQo&usp=sharing)