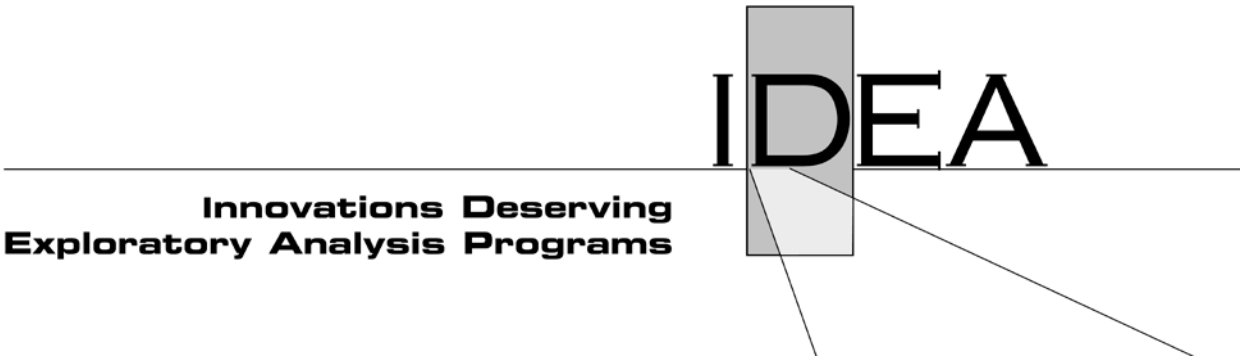


NCHRP IDEA Program

An Automated System for Pedestrian Facility Data Collection from Aerial Images

Final Report for
NCHRP IDEA Project 209

Prepared by:
Dr. Yuanyuan Zhang
Dr. Chaoyang Zhang
Joseph Luttrell
University of Southern Mississippi

December 2021

TRB TRANSPORTATION RESEARCH BOARD
The National Academies of
SCIENCES · ENGINEERING · MEDICINE

Innovations Deserving Exploratory Analysis (IDEA) Programs Managed by the Transportation Research Board

This IDEA project was funded by the NCHRP IDEA Program.

The TRB currently manages the following three IDEA programs:

- The NCHRP IDEA Program, which focuses on advances in the design, construction, and maintenance of highway systems, is funded by American Association of State Highway and Transportation Officials (AASHTO) as part of the National Cooperative Highway Research Program (NCHRP).
- The Safety IDEA Program currently focuses on innovative approaches for improving railroad safety or performance. The program is currently funded by the Federal Railroad Administration (FRA). The program was previously jointly funded by the Federal Motor Carrier Safety Administration (FMCSA) and the FRA.
- The Transit IDEA Program, which supports development and testing of innovative concepts and methods for advancing transit practice, is funded by the Federal Transit Administration (FTA) as part of the Transit Cooperative Research Program (TCRP).

Management of the three IDEA programs is coordinated to promote the development and testing of innovative concepts, methods, and technologies.

For information on the IDEA programs, check the IDEA website (www.trb.org/idea). For questions, contact the IDEA programs office by telephone at (202) 334-3310.

IDEA Programs
Transportation Research Board
500 Fifth Street, NW
Washington, DC 20001

An Automated System for Pedestrian Facility Data Collection from Aerial Images

IDEA Program Final Report

NCHRP-209

Prepared for the IDEA Program

Transportation Research Board

The National Academies

Principal Investigator: Dr. Yuanyuan Zhang

Co-Principal Investigator: Dr. Chaoyang Zhang

Student Researcher: Joseph Luttrell

University of Southern Mississippi

December 5, 2021

ACKOWLEDGEMENTS

The authors express appreciation to the following agencies and individuals who funded and contributed to the study.

Rachel Carpenter and Jessica Downing, for the supervision of this project and assistance in connecting with professionals in the Department of Transportation of several states.

Dr. Robert Schneider, for the professional consulting service

Dr. Laura Sandt, Dr. Frank Proulx, Dr. Shawn Turner, and Dr. Offer Grembek, for being on the expert panel for the project, attending progress presentations, and reviewing the draft report

Dr. Inam Jawed, for the guidance and comments on project management

Table of Contents

EXECTUTIVE SUMMARY

Recognizing the importance of pedestrian facility data to safety, thirty-seven State Departments of Transportation (DOTs) have prioritized improving pedestrian facility inventory, particularly concerning crosswalks and sidewalks, as an important action item in their Strategic Highway Safety Plans. However, such information is not widely available at the state level, with only 11 states having reported collection of such data. The hesitation of conducting this complex and repetitive yet essential data collection task could be raised by the challenges inherent in the current manual or semi manual data collection approaches, including human errors, high cost for time and labor, safety concerns for data collectors, and the corresponding concerns about standardizing, updating, and maintaining data. To address this urgent need for data and the challenges in the current data collection methods, this project developed an innovative system using convolutional neural networks, an advanced machine learning method that uses deep learning, to process image data. These networks were used to automatically collect major pedestrian facility data, including sidewalk presence, crosswalk presence, and crosswalk length, from aerial images. Results of the project are summarized as follows.

1) A data acquisition workflow was developed to automatically prepare labeled sample data for sidewalk and crosswalk detection. This process automatically generated large image datasets from any given area with images tagged as “having crosswalks”, “not having crosswalks”, “having sidewalks”, and “not having sidewalks”.

2) A convolutional neural network (CNN) model was used to automatically detect and classify images into one of four classes (crosswalk, no-crosswalk, sidewalk, or no-sidewalk). These models were tested for performing crosswalk detection and sidewalk detection with an accuracy rate of 98.43% and 92.87%, respectively. These testing results demonstrate the high efficiency of automatically collecting the data with zero cost (not including the cost for tool development).

3) Innovatively, to overcome situations where sidewalks or crosswalks are occluded in the aerial imagery, a dual-perspective method was developed to double check the ground truth information for target objects by making use of both aerial and street-view images simultaneously. A test on the dataset with heavily occluded aerial crosswalk imagery showed that this method can identify most occluded images and increase detection accuracy by 49%.

4) The crosswalk mensuration model was developed to automatically obtain measurements of crosswalk length and width by identifying a bounding box that contains all the pixels that belong to a crosswalk in a given image. In addition, the coordinates of the center of the bounding box were generated and recorded as the location of the detected crosswalk.

5) A system that integrates all trained models was developed and tested using satellite imagery of Forrest County, MS provided by the Mississippi DOT. The system is presented through a graphical user interface that gives users the ability to choose when and how to run each of the models on any images that they choose. In a test using Forrest County images, 400 images were extracted and processed by the system, resulting in an accuracy as high as 99.23% for crosswalk detection, 91.26% for sidewalk detection, and 93.7% for crosswalk length mensuration. For the 466.31 square miles covered by this imagery, it was estimated that the maximum time needed to process an area of similar size would be approximately 8 days on a windows 10 computer equipped with an Nvidia GTX 1070 graphics card.

1 IDEA PRODUCT

Collecting pedestrian facility data, such as crosswalk and sidewalk presence data, on a large scale holds a number of keys to improving safety and convenience for pedestrians. Such information is necessary for finding the causing factors of pedestrian related crashes, identifying locations that would benefit from additional crosswalks or sidewalks, and evaluating the connectivity of the pedestrian network (*1*). Recognizing the importance of pedestrian facility data to safety, thirty-seven State Departments of Transportation (DOTs) have prioritized improving pedestrian facility inventory, particularly concerning crosswalks and sidewalks, as an important action item in their Strategic Highway Safety Plans. In a recently published guidebook on measuring multimodal network connectivity (*1*), it is emphasized that ‘‘results (of pedestrian network analysis) are only informative to the extent that they measure the ‘right’ network—the one that pedestrians are likely to use in real life.’’ This “right” network is composed of crosswalks and sidewalks that are present in the real world. However, crosswalk presence data is often not available on a large scale. A National Cooperative Highway Research Program (NCHRP) Synthesis on the availability of pedestrian infrastructure data (*2*) concluded that only 12 of the 31 states have the data available to the public. Only 11 states reported collection of crosswalk presence data.

The limited availability of pedestrian facility data at scale could be caused mainly by challenges inherent in the current data collection approaches, including field data collection and manual digitization based on aerial images (*3–5*). In the first approach, data collectors go out in the field to observe and measure the facilities. They record the measurements either on paper or on hand devices for future digitization. The second approach is more advanced since data collection would be conducted mainly on computers using aerial images and video logs. However, it still needs additional field investigation for ground truth verification. Human errors, high cost for time and labor, safety concerns for data collectors, and the corresponding issues related to standardizing, updating, and maintaining data could raise the hesitation of agencies to perform large-scale collection of this complex yet essential data. To address these data collection challenges, this project developed an innovative system to apply advanced machine learning techniques, such as image processing and deep learning, to automatically collect major pedestrian facility data from aerial images. To ensure the feasibility of this pioneering system and to maintain the completeness of the pedestrian network, the system collects three types of information referred to as "major pedestrian facilities", including paved sidewalk presence, marked crosswalk presence, and marked crosswalk length. Aerial images of candidate locations are the input of the system where major pedestrian facilities are detected and measured using deep learning models combined with conventional image processing techniques. Figure 1 illustrates the framework of the product. Four core models integrated in the system were developed to prepare model training data, train the facility detection model, check ground truth when street view data are available, and measure the length of the detected crosswalk, respectively. Final outputs are then stored for future integration into the existing roadway inventories owned by state DOTs.

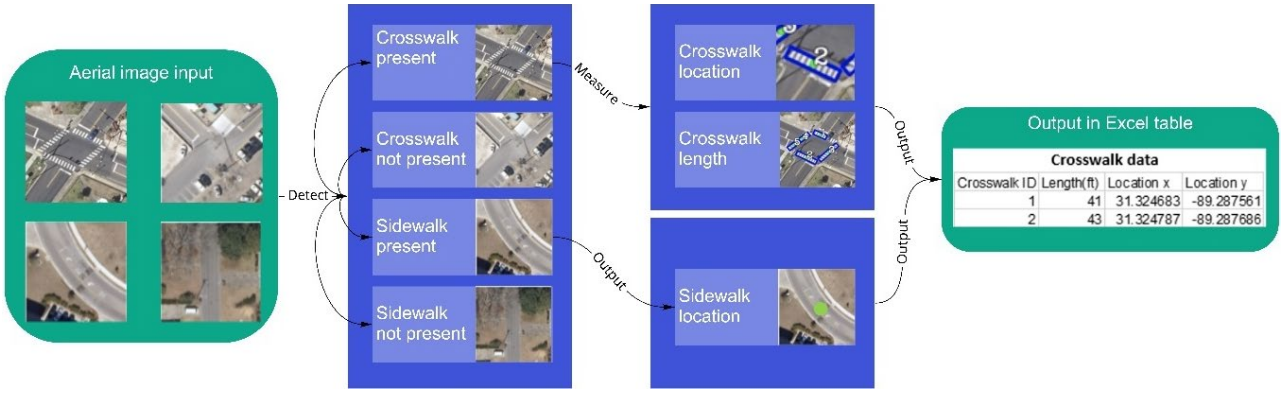


Figure 1: The product - automated data collection system, in a picture

2 CONCEPT AND INNOVATION

Promising automated methods have been studied by researchers via employing computer science techniques to automatically collect crosswalk or sidewalk presence data from aerial view or street side view images (see Table 1). Aerial view imagery includes images taken from an airplane, drone, or satellite, and provides pictures of the area from an overhead angle. Street side view imagery is taken on the street by cameras mounted on a vehicle (e.g., Google Street View and Bing Streetside View) or by cell phones. The idea of these automated methods is that, by using image processing algorithms, crosswalk or sidewalk presence can be detected automatically from images of locations of interest, even if some images are occluded. To automate the detection of these data, a traditional approach is simple image processing, such as edge detection (*6*) and image segmentation (*7*). A more advanced approach is based on machine learning technology, such as using processed images to train a Random Forest Classifier (RFC) (*8*) or Support Vector Machine (SVM) (*9*) to classify objects. Most recent improvements to the automated approach are based on deep learning models that train an Artificial Neural Network (ANN) to detect crosswalks or sidewalks automatically from street level images (*10*) and/or aerial images (*11*). There are several important limitations in the existing studies as summarized in Table 1. First, the sizes of the datasets used for training and evaluation are all relatively small and local, except in (*10*). Second, the accuracy of all the approaches is limited by the lack of a ground-truth-checking mechanism for the occluded objects. Third, the data sources used in the existing studies are either aerial or street view images, but not combined data. Rich information buried in each of the data sources has not been fully used to augment each other. Last but not least, all of the previous studies focus on one single facility, either crosswalk or sidewalk, and apply the approach in the realm of computer science or public health. There have been no studies conducted to explore the opportunity to detect the two facilities in one system and then collect related data to complete a pedestrian network for transportation agencies.

Table 1: Summary of Related Studies on Automated Crosswalk or Sidewalk Detection

Image Types	Researchers	Year	Objects Detected	Methods	Accuracy rates	Limitations
Aerial Images	Riveiro et al. (*12*)	2015	Zebra crossings	Image segmentation	83.33%	- Small training dataset size - No ground truth verification - Traditional algorithm
	Senlet et al. (*7*)	2012	Sidewalks covered by trees	Image segmentation	87%	- Occluded sidewalks only; - No ground truth verification; - Traditional algorithm;
	Mattyus et al. (*11*)	2016	Sidewalks	Semantic segmentation	< 69.8%	- Individual roads, not network; - No ground truth verification;
	Luo et al. (*13*)	2018	Sidewalks	Human aided machine learning	< 60%	- Small training dataset size; - No ground truth verification; - Traditional algorithm
Street side images	Wang et al. (*14*)	2014	Crosswalks	Support Vector Machine classifier	78.90%	- Small training dataset size - No ground truth verification - Traditional algorithm
	Poggi et al (*15*)	2015	Crosswalks	Convolutional Neural Network	88.97%	- Small training dataset size - No ground truth verification
	Ahmetovic et al. (*16*)	2016	Zebra crossings	Image segmentation	93%	- Individual roads (not a network) - No ground truth verification
	R.F. Berriel et al (*10*)	2017	Crosswalks	Convolutional Neural Network	94.12%	- No ground truth verification

To address the abovementioned limitations, the system developed here made full use of the available crowdsourcing database (e.g., Open Street Maps) and satellite images (e.g., Google

Maps) to acquire a large size dataset from different regions in the United States for model training, validation, and evaluation. Simultaneously, street level images (e.g., Google Street View) were used innovatively to solve the occluded object detection problem. Based on the literature review, this technique is the first to use aerial images, street view, and crowdsourced satellite images at the same time to intensively augment and mine rich information in aerial images. In addition, the most advanced technology of deep learning was used and combined with classical image processing techniques to guarantee improved accuracy, efficiency, and reliability of the proposed system. Because of these advantages, the resulting system is novel and useful, thus making it a significant contribution for not only the practice community but also the research community.

By using a fully automated data collection method consisting of computer object detection and classification, human involvement will be eliminated as much as possible to effectively minimize human errors. As will be discussed in the next section, the application of the advanced image processing, machine learning, and big data fusion techniques enables higher data accuracy to be achieved, compared to existing approaches used in practice and other automated methods. By automatically collecting data in-office from aerial images, time and labor cost will be largely reduced due to saved trips and human observations. Once the aerial images are updated, the data collection can be repeated easily to maintain the timeliness of the information. A state agency using this system on the updated aerial images of its own state will be able to guarantee the timeliness and consistency of the data collected in a large-scale area.

3 INVESTIGATION

The investigative approach is divided into two stages: Stage I (functional model development), and Stage II (system development and testing). In Stage I, four functional models were developed to automatically acquire labeled aerial images containing or not containing pedestrian facilities, train the facility detection model, check the ground truth for occluded facilities, and measure the numerical features of crosswalks. In Stage II, the system was developed by connecting and integrating the four models. Performance evaluation was conducted to assess the accuracy of the detected data and the efficiency of the system. Stage II also involved testing with aerial images from Mississippi Department of Transportation (MDOT) to collect future adjustments. An overview of the project tasks is shown in Figure 2.

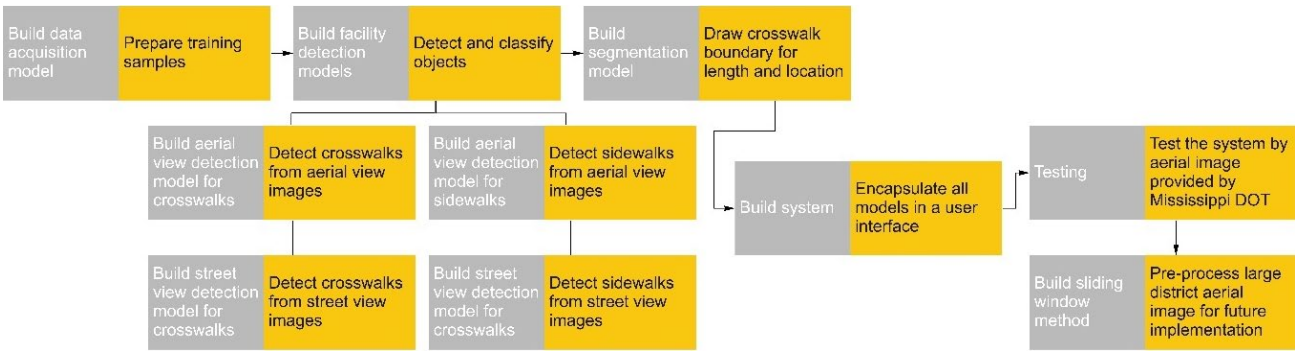


Figure 2: An overview of the project tasks.

Six tasks were conducted as explained below.

Stage I – Functional Models Development

Task 1. Develop "Sample Data Acquisition Model". In order to train a computer model to detect facilities, the first step was to prepare sample data which were labeled as "having a sidewalk" or "not having a sidewalk" for example. Since there was no ready-to-use training data for training a pedestrian facility detector, the research team developed a data acquisition model which can automatically retrieve sidewalk and crosswalk locations (as well as locations without any target facility) from Open Street Maps and then download corresponding satellite images of those locations from Bing Maps. As a result, a dataset of aerial images labeled as "sidewalk", "crosswalk", or "none" was collected and divided for training and validation.

Task 2. Train "Facility (Sidewalk and Crosswalk) Detection Model". To collect data on a large-scale efficiently, the proposed system used a deep-learning-based model developed with a Convolutional Neural Network. Before training, augmentation of downloaded sample data was required to generate input data in uniform scale and quality. Different configurations of the model were evaluated and compared using the same sample data. Finally, the most efficient one was be chosen to detect if a facility is "present", "absent", or "occluded" in a given aerial image.

Task 3. Develop "Occluded-Facility-Checking Model". Collecting pedestrian facility data purely from aerial images is a challenging task, since sidewalks in aerial images are typically highly occluded by trees and their shadows. There can also be several objects on maps that have a similar appearance to sidewalks. The "Facility Detection Model" was designed to initially detect visible sidewalks and crosswalks. For partially occluded facilities, the occluded areas are inferred with multiple priors, which fuse knowledge about roads, occluding objects, and facility structures. For completely occluded facilities, the ground truth is verified in Bing Street View

images. In this method, nearby street view panoramas that might contain the facility are acquired automatically and then used for facility verification.

Task 4. Develop “Mensuration Model”. Once a facility is detected as a “marked crosswalk”, the length of the crosswalk edge is measured automatically from the aerial image using computer image processing techniques. For visible crosswalks, the longest edge is used to measure the distance. For occluded crosswalks, assumptions are made by fusing the research team’s knowledge about road and crosswalk structure. This numerical information is then output as a feature of the crosswalk.

Stage II – System Development and Testing

Task 5. System building and evaluation. This task first completed the system by connecting and integrating the abovementioned functional models using the Python programming language and executing them on a Windows 10 desktop computer with a Nvidia GTX 1070 graphics card and an Intel i7 6700k processor. Then, pedestrian facility data collected by MDOT was pre-processed for the evaluation. Comparisons between the data collected by the proposed system and records collected manually from MDOT and Caltrans data were conducted to assess the performance of the system in terms of data accuracy, system generalization, and processing efficiency.

Task 6. Testing. The research team used the aerial images from MDOT to test the system’s generalization abilities. Since the aerial images are divided by county or city boundary, target areas that contain potential facilities along roadways or at intersections were extracted first. Layers of road center lines and intersection points from state DOTs were used as the reference layers for area extraction. Extracted images were then pre-processed to meet the system requirements and fed to the system as an input. The entire process was conducted with the advisory team comprised of Caltrans and MDOT and their feedback and suggestions were acquired. A summary of potential adjustments was documented for future improvement and implementation.

3.1 DEVELOP “SAMPLE DATA ACQUISITION MODEL”

The sample data acquisition model was designed to prepare tagged images of crosswalks and sidewalks for the purpose of training the detection model. To realize this function, this model was designed as a pipeline of scripts that combines crowdsourced tags of pedestrian facility locations from OpenStreetMap (OSM) with their corresponding satellite images from Bing maps.

As a result, the data acquisition model produced a large number of images (tagged as "crosswalk" or "no-crosswalk") for use in training and testing the deep learning models used in the project. Figure 3 illustrates this sample data acquisition process and the different processes used to gather positive and negative samples. First, a bounding box (depicted in Figure 3 with a red rectangle) is defined and passed into the data acquisition program in the form of a pair of latitude and longitude coordinates for the lower-left and upper-right corners of the region. Then, locations within this bounding box corresponding to OSM tagged crosswalks (blue markers in Figure 3) are sent to the Bing Maps RESTImagery API in order to obtain aerial images of crosswalks (positive samples). Using these known crosswalk locations, the Bing Maps RESTRoutes API is used to calculate a route between a given pair of crosswalks and determine a number of points on the route which do not contain OSM tagged crosswalks (yellow markers in Figure 3). Rather than randomly selecting points within the input bounding box region (which may produce images of forests, bodies of water, and other undesired scenes), using this route-based method ensures that the negative samples will be images of roads.

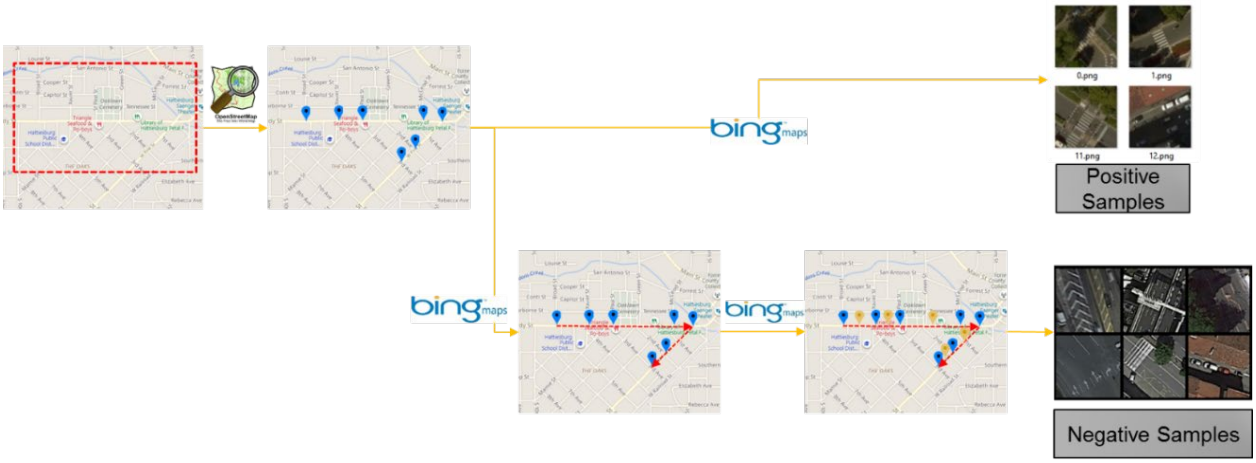


Figure 3: Sample data acquisition procedure for generating training and testing samples for crosswalk detection.

In order to prepare for the development of the facility detection models, this data acquisition strategy was used to download an initial set of aerial images of crosswalks from Bing Maps using OSM coordinates. A total of 4,940 satellite images tagged as "crosswalk" and 10,937 satellite images tagged as "no-crosswalk" were generated automatically for a randomly

selected region on Bing Maps. Figure 4 shows a workflow diagram of the procedure and an example of its output for this initial dataset. Filtering is done to ensure that these "no-crosswalk" points are not too close to crosswalk points (to minimize the chance of generating mislabeled images) before using the RESTImagery API again to download aerial images (negative samples). After this basic filtering to remove exact duplicate images (resulting from html request errors), the number of crosswalk images remained the same, but the number of no-crosswalk images decreased to 10,299.

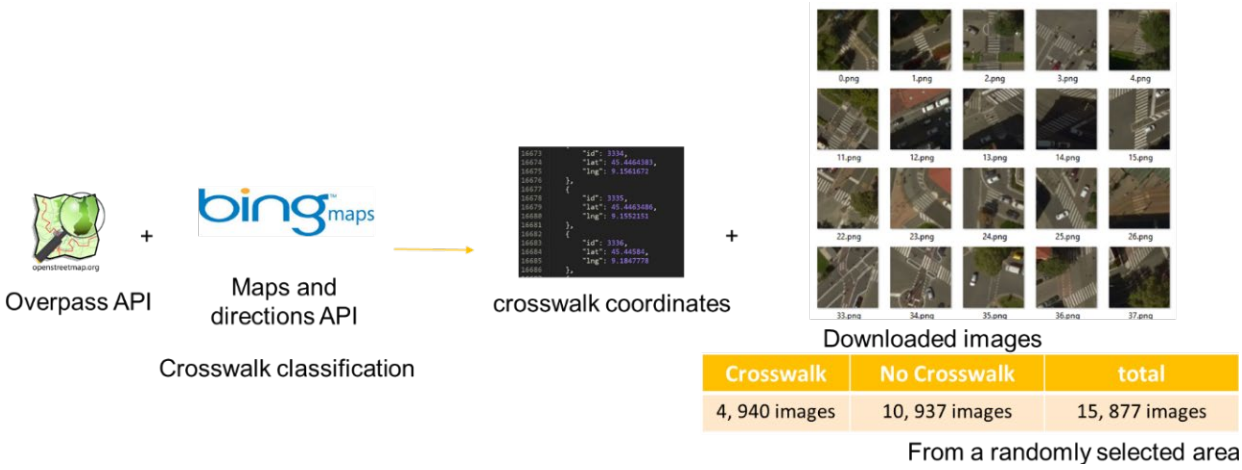


Figure 4: The data acquisition program used to obtain 15, 877 tagged images from a randomly selected area.

3.2 TRAIN "FACILITY DETECTION MODEL"

After this dataset was gathered, the development of the facility detection model began. This model is responsible for detecting the following object categories: 1) marked crosswalks in satellite-view imagery, 2) marked crosswalks in street-view imagery, 3) paved sidewalks in satellite-view imagery, and 4) paved sidewalks in street-view imagery. These categories were approached as independent tasks, and four sub-models were created to handle each one.

Detect Crosswalks and Sidewalks from Satellite Images

In order to begin the development process for these sub-models, it was necessary to process the images downloaded by the "Sample Data Acquisition Model" into a form that was usable for deep learning. This involved further duplicate filtering (coordinate-based) and random splitting into train/test/validation sets. The applied duplicate filtering technique works by converting the location (latitude and longitude) of each coordinate pair (point) to Universal Transverse Mercator

(UTM) coordinates and using those results to calculate the Euclidean distance between each point and every other point in the dataset. Table 2 shows the results (before distance filtering, after distance filtering) of this filtering process on the satellite-view crosswalk dataset and the distribution of the images after they were randomly split into training, testing, and validation subsets (70%, 20%, and 10% respectively). The model that produced these results is a convolutional neural network (CNN) that is based on the VGG16 architecture with weights pre-trained on the Imagenet dataset (available in the Keras "applications" API). This model was modified to replace the output layer with two units (instead of 1,000) for performing binary classification, and all layers remained unfrozen for the entire duration of training (200 epochs). The learning rate used here was 0.0001, and the batch size was 16.

Table 2: The number of images in the satellite-view crosswalk dataset

Subset	Positive (crosswalk)		Negative (no-crosswalk)		Total	
	Before distance filtering	After distance filtering	Before distance filtering	After distance filtering	Before distance filtering	After distance filtering
Training	3458	1467	7209	1599	10667	3066
Validation	494	210	1030	228	1524	438
Test	988	419	2060	457	3048	876
Total	**4940**	**2096**	**10299**	**2284**	**15239**	**4380**

Using these processed images to train the satellite-view crosswalk detection model, a 98.43% test accuracy was obtained (before performing coordinate-based distance filtering). After the aforementioned filtering process, the same model trained on the distance filtered dataset shown in Table 2 achieved a 96.35% accuracy as shown in Figure 5 (A) despite the considerable amount of overlapping images that were filtered out. This was done to ensure that it was still possible to achieve good test accuracy without the potential bias caused by images with duplicated regions of pixels existing between the training and test sets. The accuracy scores were based on the method shown in Formula 1.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Where:

TP – The number of images correctly predicted to contain the positive class (crosswalk detected as crosswalk);

TN – The number of images correctly predicted to belong to the negative class (non-crosswalk detected as non-crosswalk);

FP – The number of images incorrectly predicted to contain the positive class (non-crosswalk detected as crosswalk);

FN – The number of images incorrectly predicted to belong to the negative class (crosswalk detected as non-crosswalk).

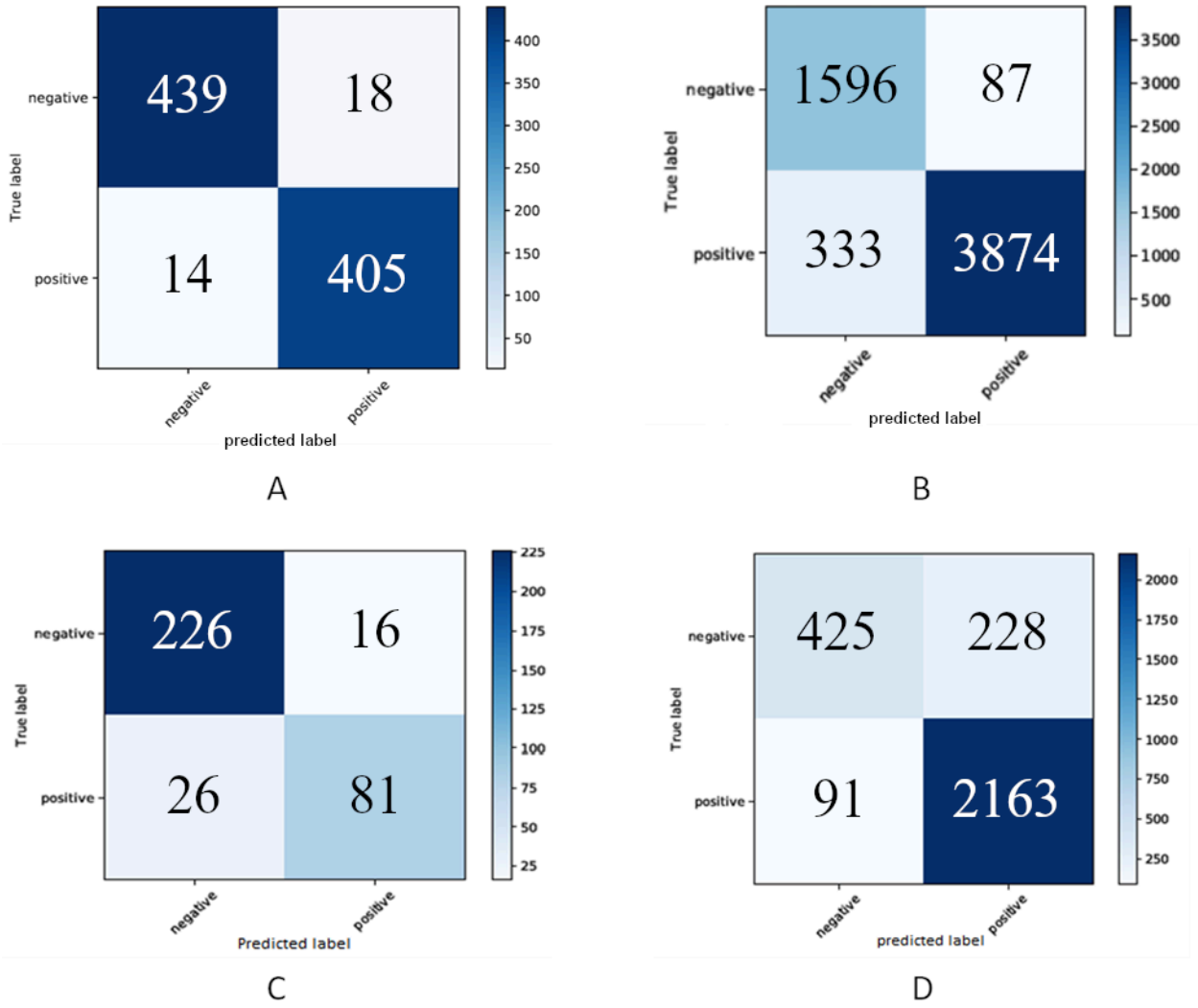


Figure 5: The confusion matrices for the separate components of the facility detection model. (A) satellite-view crosswalk model. (B) satellite-view sidewalk model. (C) street-view crosswalk model. (D) street-view sidewalk model. For each matrix, the predicted labels for the images are on the x-axis (positive = presence of facility, negative = absence of facility) and their true labels are on the y-axis.

After the crosswalk satellite detection model was finished, data was gathered to begin the training of the sidewalk detection model. This was done using a slightly modified version of the data acquisition procedure. The major difference between collecting crosswalk data and sidewalk data is that locations are queried using different tags in OpenStreetMap (OSM). For sidewalk points, this involves using sidewalk=left, sidewalk=right, or sidewalk=both. Also, instead of generating negative samples by following a routePath between crosswalks (as performed in the data acquisition for the crosswalk data), OSM nodes with the tag sidewalk=none are directly

requested. A new area (separate from the area used in the crosswalk dataset) was selected and all the previously mentioned data acquisition and filtering techniques were applied. This produced a new sidewalk facility detection dataset listed in Table 3. Then, a new CNN model (similar to the one used in the satellite-view crosswalk facility detection model) was trained to use this sidewalk dataset. Figure 5 (B) shows that the resulting satellite-view sidewalk detection model was able to obtain a 92.87% test accuracy (defined by formula 1).

Table 3: The number of images in the satellite-view sidewalk dataset organized by class (sidewalk vs no-sidewalk) and the subset they belong to (training, validation, or test)

Subset	Positive (sidewalk)	Negative (no-sidewalk)	Total
Training	14723	5892	20615
Validation	2103	842	2945
Test	4207	1683	5890
Total	**21033**	**8417**	**29450**

Detect Crosswalks and Sidewalks from Street-View Images

The next two models were trained using street-view images and the same architecture as the satellite-view models. These models play an important role in both the facility detection model and the occluded facility detection model. Once again, the data acquisition process was adjusted for obtaining and filtering street-view images. This was done by switching to the Bing streetside imagery API and implementing a new method for obtaining images from each OSM location. The most important technique for obtaining street-level images from a given OSM location is calculating heading. This is due to the fact that the extracted OSM locations (such as a node tagged as having a crosswalk) are represented by a single point (latitude, longitude). However, the street-level imagery is stored as a panorama which can generate many possible images that are the input size that the network uses. To solve this problem, it is first necessary to query for the image metadata of the street-view panorama closest to the OSM point of interest. This gives an image which is often too close to the facility of interest or directly on top of it. Furthermore, since the default query simply faces the camera north, the returned portion of the panorama will likely not contain the point of interest. Before calculating the heading necessary to solve this problem, a query is formed to retrieve a new point 10 meters (an empirically determined

distance) away. This is done using the reverse heading of the current image's value for heading in the metadata to approximate moving backwards or forward along the road depending on the direction the imagery capturing vehicle was moving. The heading was calculated based on Formulas 2 – 4.

$$Heading = atan2(X, Y) \tag{2}$$

$$X = \cos\theta_b * \sin\Delta L \tag{3}$$

$$Y = \cos\theta_a * \sin\theta_b - \sin\theta_a * \cos\theta_b * \cos\Delta L \tag{4}$$

Where:

θ_a – The latitude (in degree) of the original point of interest

θ_b – The latitude (in degree) of the new point that is 10 meters away

ΔL – The difference in longitude between the two points

Querying once again using the new point 10 meters away and turning the camera using the newly calculated heading has a much higher chance of producing a reliable image with the point of interest in the frame. This enabled the creation of a street-view crosswalk dataset and a street-view sidewalk dataset (both sampled from arbitrarily chosen locations that are geographically separate). Table 4 and Table 5 give the size and class distribution for these new street view image datasets.

Table 4: The number of images in the street-view sidewalk dataset organized by class (sidewalk vs no-sidewalk) and the subset they belong to (training, validation, test).

Subset	Positive (sidewalk)	Negative (no-sidewalk)	Total
Training	7889	2288	10177
Validation	1127	327	1454
Test	2254	653	2907
Total	**11270**	**3268**	**14538**

Table 5: The number of images in the street-view crosswalk dataset organized by class (crosswalk vs no-crosswalk) and the subset they belong to (training, validation, test).

Subset	Positive (crosswalk)	Negative (no-crosswalk)	Total
Training	374	847	1221
Validation	53	121	174
Test	107	242	349
Total	**534**	**1210**	**1744**

3.3 DEVELOP "OCCLUDED-FACILITY-CHECKING MODEL"

In order to create the occluded facility checking model, it was first necessary to obtain a dataset of images with occluded regions manually verified by humans. To accomplish this, an area specifically chosen for its highly occluded satellite imagery was used to create an "external test" dataset. This dataset served to both provide a geographically separate test for the models (since the "local" test dataset for each model is sampled from locations in the same city that the model is trained on) and to provide examples of challenging occluded satellite imagery. After using the data acquisition model to obtain these images, the manual verification interface was used to verify the label (crosswalk or no-crosswalk) of each image and also to note if the image is occluded or not. Figure 6 shows an example of the verification interface in operation. This interface first opens the image file for each location and asks the human verification worker if any occlusion is present. The interface itself is running in the command prompt on the left and displays choices for the user to manually label each location. The two images on the right both correspond to different views of the same location. The user can monitor the street view on the right monitor to verify the presence of the pedestrian facility in the aerial image if necessary. The verification system keeps a log of each entry including a timestamp and can be paused and resumed at any time. After checking for occlusion and confirming the class label of the image, the manual verification script also presents several options to the human verification worker. This includes the ability to delete the image, move it to the opposite class (switch the label), open a browser tab with Bing maps showing the location on the map, or quit and save their progress.

At the end of each session, the results of the evaluation are stored as a json file which records the user's decisions for each image and allows the session to easily be resumed later.

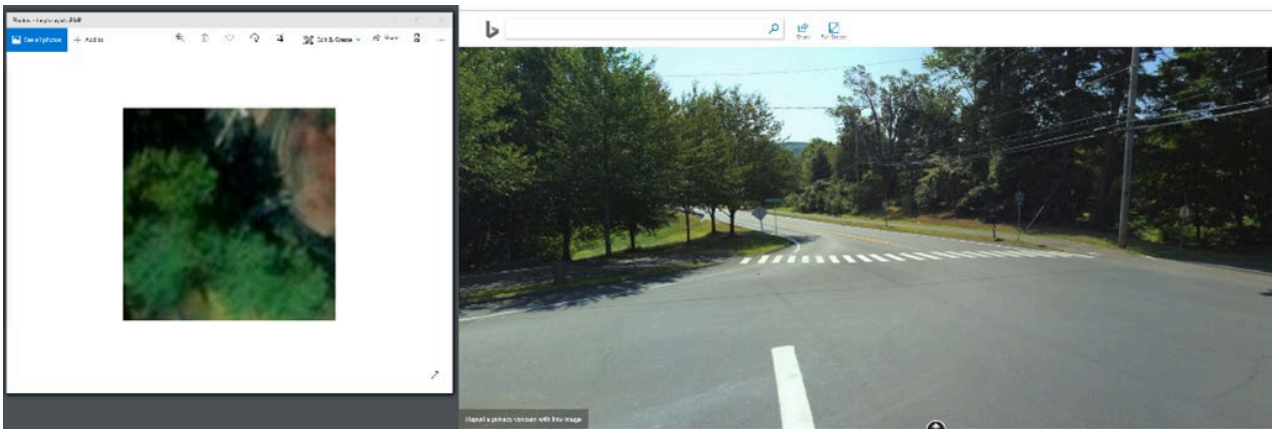


Figure 6: The python verification interface running on a desktop pc with two monitors.

Table 6 describes the distribution of this manually verified external test dataset and statistics about the amount of occlusion in the data. Unlike the other datasets, it was not split into training, validation, and testing subsets since it is only used as a test.

Table 6: Manually verified external test dataset size (number of images per class) and occlusion rates (percentage of manually verified images that are occluded).

Classes	Number of images	Aerial occlusion	Street side occlusion
Crosswalk	344	54.07%	24.42%
Non-crosswalk	345	60.76%	32.17%
Total	**689**	**57.41%**	**28.30%**

All the crosswalk models were then evaluated and their results on both the local and external test datasets were listed in Table 7 in terms of accuracy, precision, and recall. Accuracy is a measure used to quantify what proportion of images was correctly detected. In addition, precision and recall are the other two measures to evaluate the performance of a machine learning model. Precision attempts to answer what proportion of positive identifications was actually correct. For example, if the precision of the model to detect crosswalks from an un-occluded satellite view is 98.45%, it means that when it predicts a crosswalk is a crosswalk, it is correct 98.45% of the time. Recall quantifies what proportion of actual positives was identified correctly. Therefore, if the facility detection model's recall for detecting crosswalks from an un-

occluded satellite view is 96.66%, it means that it correctly identifies 96.66% of all real crosswalks. Here, accuracy was defined by Formula 1, while precision and recall were defined by Formula 5 and 6 as shown in below.

$$Precision = \frac{\text{TP}}{TP + FP} \tag{5}$$

$$Recall = \frac{\text{TP}}{TP + FN} \tag{6}$$

Where:

TP – The number of images correctly predicted to contain the positive class (crosswalk detected as crosswalk);

FP – The number of images incorrectly predicted to contain the positive class (non-crosswalk detected as crosswalk);

FN – The number of images incorrectly predicted to belong to the negative class (crosswalk detected as non-crosswalk).

Table 7: The facility detection model evaluated on both the local (local crosswalk and local sidewalk) test datasets and the external test dataset.

Dataset Classes	Accuracy	Precision	Recall
Satellite-view crosswalk (local test) without occlusion checking	98.43%	98.45%	96.66%
Street-view crosswalk (local test) without occlusion checking	84.81%	75.00%	75.70%
Satellite-view crosswalk (external test) without occlusion checking	55.59%	77.94%	15.41%
Street-view (external test) without occlusion checking	82.58%	83.73%	80.81%
Crosswalk detection (external test) with occlusion checking	83.02%	89.82%	74.42%

It is worth noting that the evaluation for the occluded facility checking model is calculated a bit differently compared to the other models, which is done using soft voting between the satellite-view crosswalk model and the street-view crosswalk model. This simply performs two predictions for each location (one at street-level and one with satellite imagery using the appropriate models). The resulting output of the final softmax activation for each model is added together and then treated as a single prediction. In other words, for each location, the class with the maximum sum of probability values between the satellite-view and street-view models is chosen as the final prediction. As shown in Figure 7, predictions from the three crosswalk models for both views of one location from the external test class probabilities were given as [% negative, % positive] where the positive class represents crosswalks and the negative class represents the absence of crosswalks. **A** is a prediction by the aerial model in which the probability for the negative class was higher due to the highly occluded nature of the crosswalk. **B** shows the same crosswalk from street level and is a prediction from the street-level model which shows a very confident and correct positive prediction. **C** is the result which was used as the occluded facility checking model's final prediction for this target and shows the soft voting combination of the two other models' predictions from **A** and **B**. The final choice is the largest sum of probability values (the positive crosswalk class here).

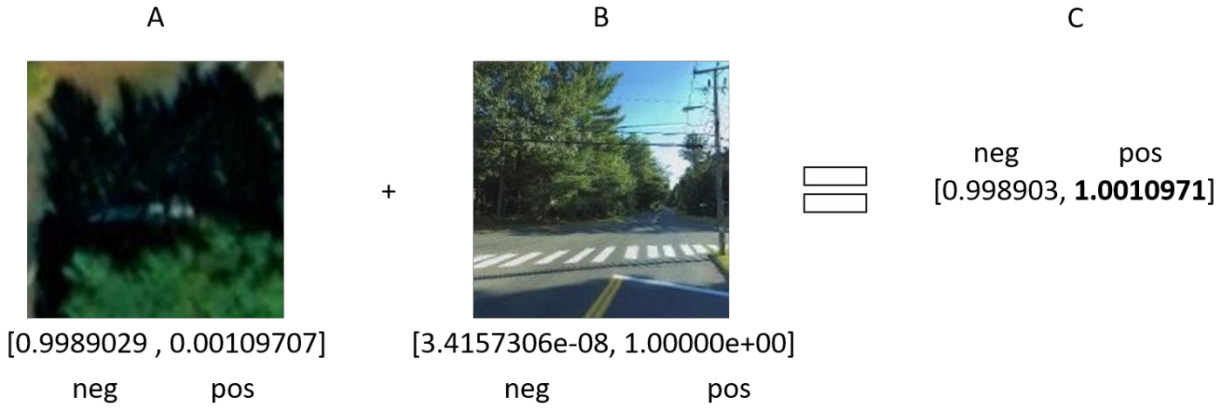


Figure 7: Predictions from the three crosswalk models for both views of one location from the external test

These measures were also compared with the occluded facility checking model evaluated for crosswalk detection on the external test dataset (also summarized in Table 7). The external test dataset has many instances of crosswalks being very out of focus or otherwise completely

occluded by trees or other obstacles (the manual verification results revealed an estimated total of 57.41% of the images in the dataset were occluded). As a result, the facility detection model suffers greatly in the evaluation when compared to the occluded facility checking model. After implementing the occluded facility check, the external test accuracy for crosswalk detection jumped from 55.59% to 83.02%. Perhaps more importantly, the recall improved from 15.41% to 74.42%. This substantial increase in recall means that a large majority of the occluded or otherwise unrecognizable crosswalks in the satellite imagery that were initially missed were able to be recovered and correctly classified by incorporating this occluded facility check.

3.4 MENSURATION MODEL

Based on a literature review of image processing techniques for road facilities, it was determined that segmentation was the best method to use for obtaining measurements of crosswalks. Using these techniques, it was possible to classify the presence or absence of crosswalks in an image while also identifying their exact location in the image. Given enough image metadata (ground resolution, etc.), the results of this segmentation can be used to obtain an accurate measurement of the target facility. Figure 8 shows an overview of the process used to develop this model. In total, 100 images (with 65 used for training) were processed and used in this manner to create the trained segmentation (mensuration) model. The highlighted yellow pixels in the training input image represent a manually drawn mask used to train the model to identify the crosswalks in the training images. The output is a blue box drawn around the predicted crosswalk region and a green dot at the center of this region. The length of the longest side of this bounding box is used as the predicted length of the crosswalk and the coordinates (latitude/longitude) of the centerpoint are used as the location of the crosswalk. A more detailed test of this model is discussed in section 3.7.

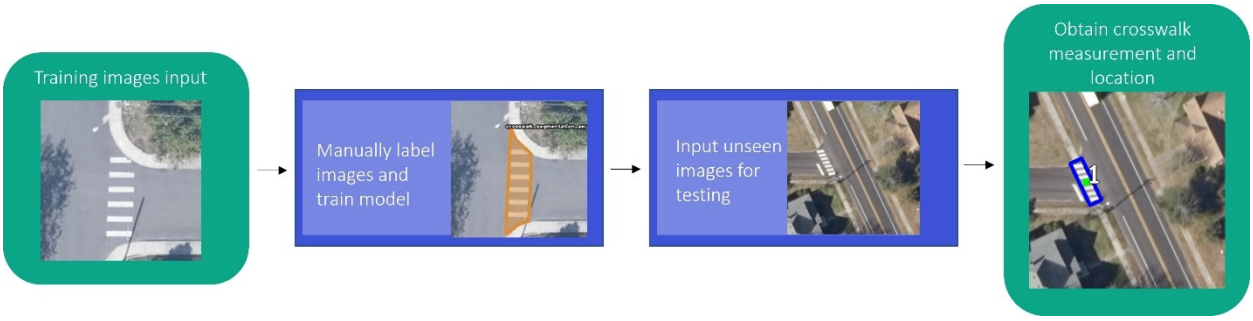


Figure 8: An overview of the development of the mensuration model.

3.5 SYSTEM DEVELOPMENT AND EVALUATION

The models developed in this project have been organized in a software package that allows users to automatically run them on their own input images. Figure 9 shows an overview of all of the developed models that are part of the system (including the ones not currently shown in the interface) and how they work together.

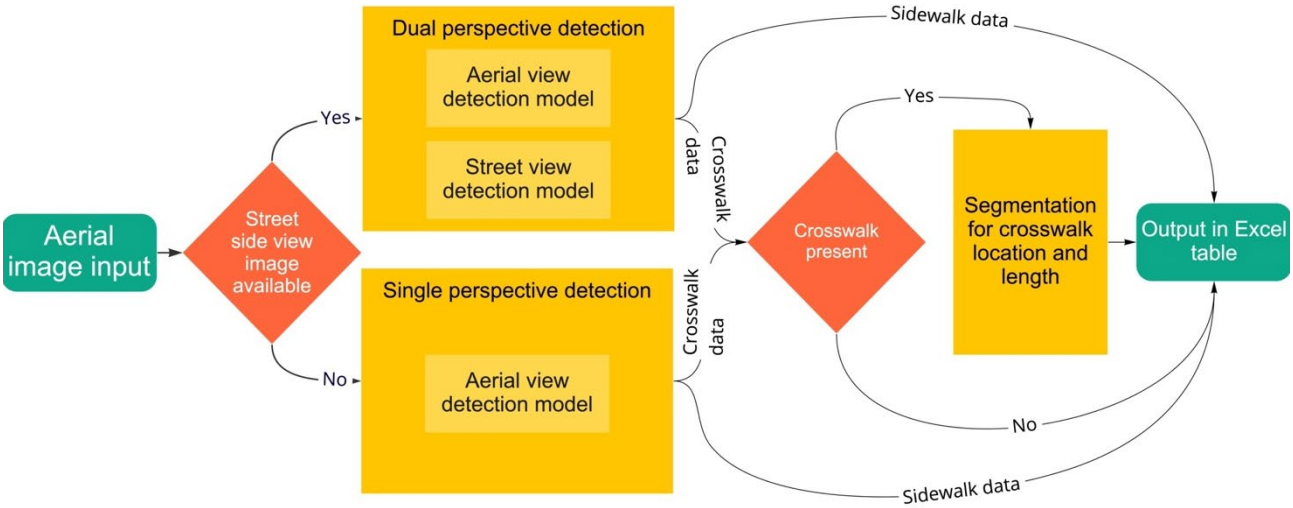


Figure 9: An overview of the components of the system.

This program organizes the scripts that control the data input, model loading, and prediction processes for all of the previously developed models into a single system that can be used with little knowledge of Python and the other technologies involved in this project. It uses a simple graphical interface developed with PySimpleGUI to allow users to automatically apply any of the models to either one image or an entire directory of images at once. Each function is programmed to run independently to save processing time if only one type of prediction is needed, but they can be combined with others to produce more detailed reports using any combination of prediction methods that the user chooses. All code files for the program and the user interface can be found in an online OneDrive folder through the link in below:

https://smttt-my.sharepoint.com/:f:/g/personal/w997046_usm_edu/EkEXWxihmUtHvejZteJDLZcBVhmLPLclGiJORDVFYXOuUg?e=tu8dBC

Figure 10 shows a screenshot of the main interface of the system and one of the options for processing aerial images. The main window (A in Figure 10) launches when the program starts and allows users to pick an operation mode based on the type of prediction they want to perform first. This example examines a single crosswalk at a time in an aerial image. The aerial crosswalk detection model is automatically loaded in addition to all of the relevant python libraries, scripts, and data necessary to perform a prediction. The interface then displays all available images in the directory chosen by the user. Upon clicking an image in the list, the system performs the prediction and displays both the input image and the result produced by the model. This mode can be used to quickly check the model's performance on a few images without committing to processing a large number of images at once.

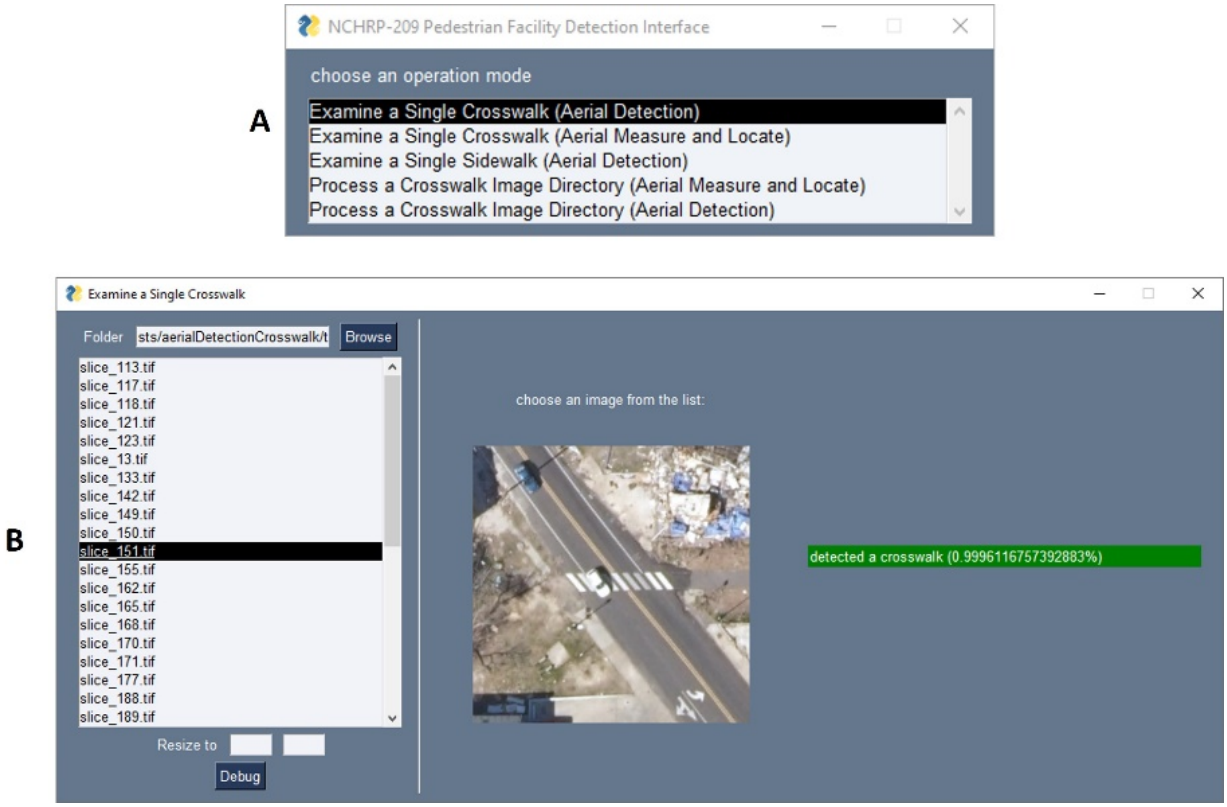


Figure 10: The interface of the facility detection system showing the main window (A) and the single crosswalk examination window (B)

A similar example using a sidewalk image instead can be seen in Figure 11. It should be noted that this interface can run any of the models produced by this project so far, but it is currently set to only use aerial images and the tested functions listed in the main window.

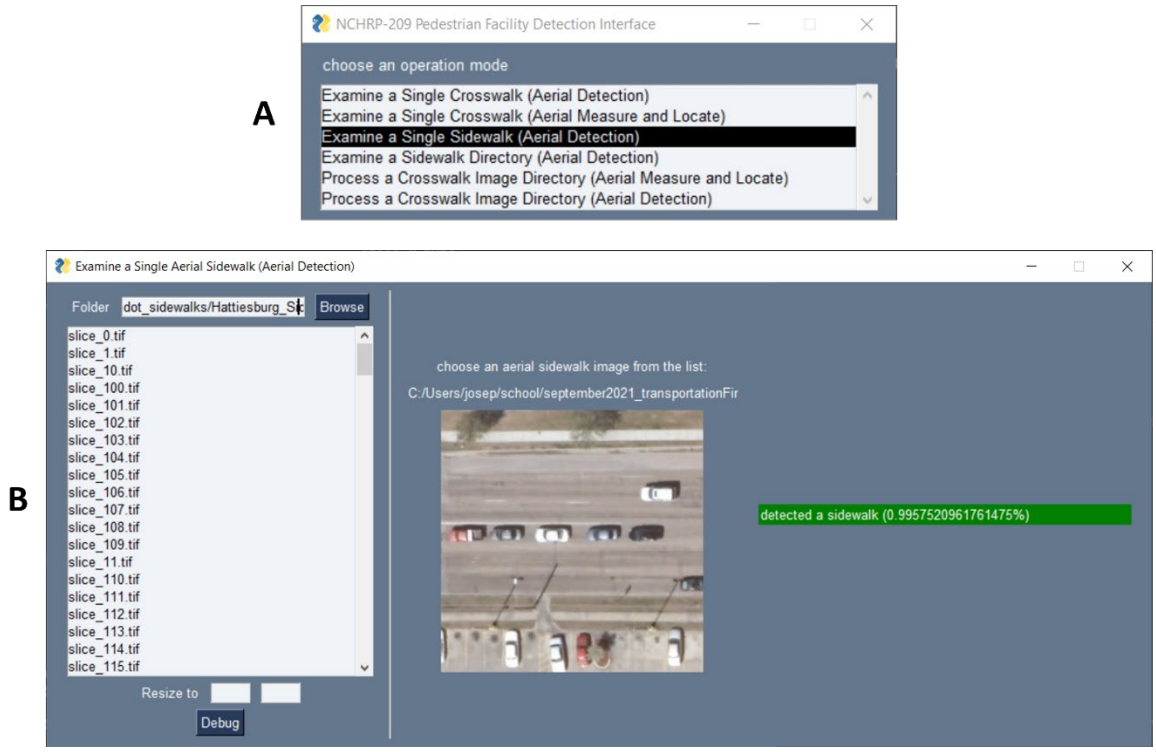


Figure 11: The interface of the facility detection system showing the main window (A) and the single sidewalk examination window (B)

3.6 TESTING

In order to test the various components of the system, a set of 400 testing images was extracted from Satellite imagery of Forrest County, MS. This data is available on the MARIS (Mississippi Automated Resource Information System) website. The images used in this test were manually extracted from this data using predefined shapefiles that were calculated to produce 256x256 pixel square images (at the resolution of 0.5 feet/pixel) that were then automatically extracted with a gdal python script (examples displayed in Figure 12). This produced georeferenced slices that were an appropriate size for the models with 200 images focused on testing crosswalk detection/mensuration and 200 images for testing sidewalk detection. The locations for the crosswalk images were extracted from intersections, and the sidewalk images were chosen from road segments between two adjacent intersections included in the crosswalk locations list. In the various tests conducted here, some of these images were removed to compensate for problems caused by images that were not optimal for processing. Due to a lack of available street view

imagery, only aerial imagery and the corresponding models were tested here in order to show an example of the performance of the system specifically for data owned by DOTs.

Figure 12: An example of two images from each testing dataset. Row A shows positive and negative crosswalk samples. Row B shows positive and negative sidewalk samples.

Therefore, tests were conducted to show the system's performance in four key areas: (1) aerial crosswalk detection, (2) aerial sidewalk detection, (3) aerial crosswalk mensuration, and (4) aerial crosswalk detection generated from segmentation results. Figure 13 shows confusion matrices for the first three of these tests, and Table 8 describes the Forrest County test datasets.

Table 8: Three Forrest County Test Datasets

Dataset	Target	Positive	Negative	Total
Forrest Test 1	Crosswalk	20	119	139
Forrest Test 2	Crosswalk	20	110	130
Forrest Test 3	Sidewalk	86	17	103

For the detection tests involving crosswalks, the number of images was reduced to 139 (Forrest Test 1) by removing positive examples (images with crosswalks) if they did not contain at least one crosswalk with a "zebra" pattern (images without any crosswalks are still retained as negative examples). The aerial crosswalk detection model (B in Figure 13) was 92.81% accurate with this data. However, by using QGIS to increase the zoom level of the images in the Forrest Test 1 dataset, it was possible to more closely match these images to the appearance of the

images that the aerial detection model was trained on. These new manually zoomed images were used as the Forrest Test 2 dataset after removing 9 duplicate images. The aerial crosswalk detection model was 99.23% accurate in this test. Also, using the Forrest Test 1 dataset with the crosswalk segmentation model and interpreting its results as classification instead of segmentation (A in Figure 13) produced an accuracy score of 92.03%. This was done by counting any detected pixels in the image as a positive prediction for the whole image and calculating accuracy normally according to Formula 1. For aerial sidewalk detection, the model was 80.5% accurate on the original 200 images, but it was then able to achieve 91.26% accuracy (C in Figure 13) using only ideal images. These ideal images (Forrest Test 3) were extracted by removing images where the sidewalk (or the side of the road in images with no sidewalk) was occluded, damaged, or otherwise not optimal for the model to process.

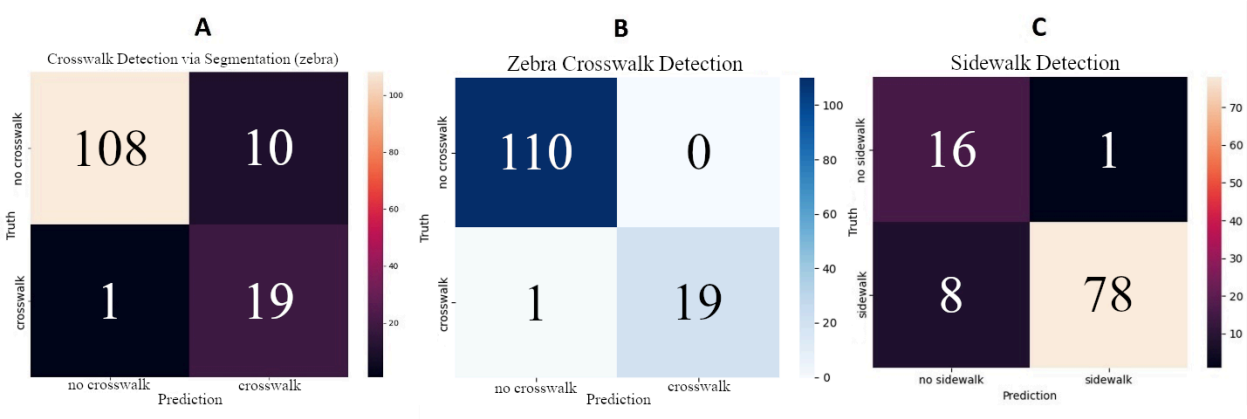


Figure 13: Confusion Matrices for the results produced by the Forrest County image testing. A is zebra crosswalk detection results interpreted from the crosswalk segmentation model (Forrest Test 1 data). B is zebra crosswalks detection using the aerial crosswalk detection model (Forrest Test 2 data). C is sidewalk detection using the aerial sidewalk detection model (Forrest Test 3 data).

Finally, the mensuration capabilities of the segmentation model were tested using the same Forrest Test 1 images used to test the aerial crosswalk detection model. In these images, it was found that there were 37 zebra crossings, and 20 of them were ideal (no occlusion or worn markings). For the full set of 37 zebra crossings, 30 of the crossings were detected with an average length measurement accuracy of 74.3% (calculated as 1 minus the percent error of the predicted and real measurement of the crossings). For the 20 ideal zebra crossings, 17 were detected (the 3 remaining crosswalks were not detected) with an average measurement accuracy of 93.7% according to the following formula where c is a zebra crosswalk, n is the total number

of zebra crosswalks that were detected by the mensuration model (17 here), y_true_c is the true length of crosswalk c (obtained from the measurement tool in Bing Maps or QGIS), and y_pred_c is the predicted length of crosswalk c.

$$Average\ Length\ Measurement\ Accuracy = 1 - \frac{\sum_{c=1}^{n} \frac{|y\_pred_c - y\_true_c|}{y\_true_c}}{n} \quad (7)$$

Figure 14 shows the process of obtaining these measurements for one input image (Figure 14 A). The first step for scoring a segmentation model is to create the ground truth by manually labeling the crosswalks in the image (done here by using an annotation program called coco annotator in Figure 14 (B). The general procedure used when manually creating these masks was to keep the drawn labels as close to the crossing as possible and to extend the labeled region all the way to the curb unless it intersects another crossing. This information was mostly used for training the models, but it also can be used for various other types of accuracy calculations during testing.

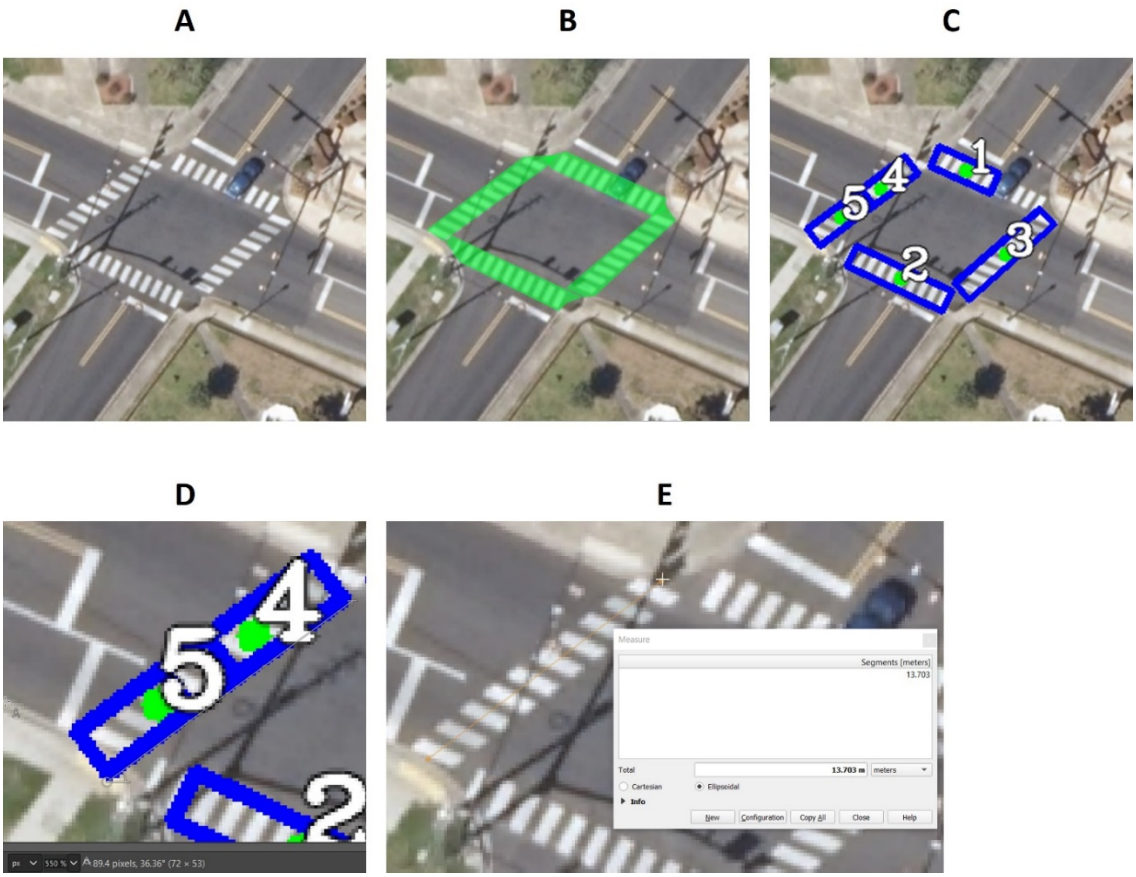


Figure 14: The process of evaluating the length measurement accuracy for an input image (A) using five steps (B-E).

However, for this test, manual measurements of the crosswalks were taken according to some simple rules that were followed when creating annotations for the original training data. As demonstrated in by the yellow line in Figure 14 (E), the measurement tool in QGIS was used to measure from curb to curb through the middle of crossing (trying to follow the angle of the crossing like the labels drawn in Figure 14 (B). Then, this value (13.703m in this case) is compared with a measurement of the results of the segmentation model (C in Figure 14) generated from the input (A in Figure 14). In Figure 14 (C), there are actually 5 separate detections for the 4 crosswalks. In this test, the pixels of multiple detections within a single crosswalk were measured as one detection, as shown in Figure 14 D. This measurement is obtained by counting the number of pixels that make up the combined length of the predicted bounding box/boxes (not counting any overlap between boxes) on a crosswalk and multiplying by the ground resolution of the image. Here, it was seen that 89.4 pixels (in Figure 14 D) multiplied by the ground resolution of 0.1524 meters/pixel gave a distance of 13.626m and a

difference of only 0.56% from the true measurement. Repeating this procedure and averaging the results produced the previously mentioned measurement accuracy scores for this test. Figure 15 shows a correctly working example where the true length (9.251m) and the predicted length (9.113m) are very close. The relatively noisy predictions in Figure 14 can likely be attributed to a number of things, such as a car blocking a portion of crossing detection 1, and a shadow intersecting the corner of the crosswalk that was detected as 4 and 5. The current segmentation model seems to be more likely to skip or split a detection if there are any obstacles in the crossing region.

Figure 15: A correctly working example where the predicted length was only about 0.13m off from the true measured length.

In this test, it was found that the facility detection models achieved accuracy scores as high as 91.26%, 93.7%, 99.23%, and 92.03% for sidewalk detection, crosswalk length mensuration, crosswalk detection, and crosswalk detection via segmentation, respectively. This shows, once again, the importance of utilizing multiple models to strengthen the system's defense against the natural variability of real-world data. In a case where the input images for one of the models are compromised by occlusion or other quality issues, the other models should be able to correct these mistakes and greatly increase the chance of 100% of images containing at least one zebra crosswalk being detected. Although there are still some false positive detections, this is preferable since it would be easier for a human user working with large satellite images to filter out false crosswalk detections than to find an undetected crosswalk that they were never informed of by the system.

3.7 TIME AND COST ESTIMATE

The amount of time and funding necessary to apply the system to a new area can be estimated using the data available for Forrest County, MS. Given the current input image size used by this system (256x256 pixels) and the total pixel size of the Forrest County satellite imagery, it is possible to calculate the following value for the maximum number of images needed to process the area from one viewpoint. Here, t_{aerial} is the currently required time for processing one of these images with the aerial view detection model (~0.2361 seconds). Given the same image, t_{street} is the time necessary to apply the street view model (0.0347 seconds) and $t_{mensuration}$ is the time needed to apply the mensuration model (0.261808 seconds). These runtime estimates were obtained using a standard Windows 10 desktop equipped with an Nvidia GTX 1070 GPU and an Intel i7-6700k CPU. The following formulas show how these numbers are used in the estimates here.

$$total\_images = \frac{81{,}900{,}000{,}00\ total\ pixels}{65{,}536\ input\ pixel\ size} = \sim 1{,}249{,}695\ images \tag{8}$$

$$total\_time = required\_images(t_{aerial} + t_{street} + t_{mensuration}) = \sim 8\ \text{days} \tag{9}$$

$$total_{cost} = required\_images(calls_{aerial} + calls_{street}) = \sim 3{,}749{,}084\ \text{API calls} \tag{10}$$

Here, the value of $total\_images$ represents all of the possible images that can be extracted from the 466.31 square miles of satellite imagery contained in this test data. Using this, Formula 9 shows how the total processing time can be estimated. Then, Formula 10 shows an estimate for the total cost of these operations in terms of maps API calls (Google and Bing both offer different rates and various options for bulk pricing when many calls are needed). Here, $calls_{aerial}$ is equal to one call and $calls_{street}$ is equal to two calls. For aerial images, one call will provide one aerial image of a location (256x256 pixel images centered at the given coordinates). For street view images, one call also provides one street view image for the same location. However, two API calls are currently needed to properly frame the crosswalk within a street view image and extract it from the full street view image. Also, it is important to note that the number of images required to process this example area can be greatly reduced by extracting

images only along roadways. Furthermore, street view imagery would not be available for areas that are not located on a road. All of the necessary processing to extract road locations from the list of all possible locations could be done with existing GIS data and would not require additional API calls.

4 PLANS FOR IMPLEMENTATION

To prepare the implementation plan and ensure the future system's compatibility with other state DOTs environments, a survey of 19 questions about aerial image quality, update frequency, image time stamps, software environment, computation capacity, etc. was designed to collect information from all state DOTs. In this way, the standard level of requirement for the system would be determined and considered for the implementation of the project results. The survey has been distributed through the Transportation Alternatives / Recreational Trails Program and the Office of Human Environment, Livability Team at the Federal Highway Administration on February 10th, 2021. As of October 14th 2021, the research team has not received the survey results yet. Since the survey is beyond the scale and tasks of this project, it can be conducted again in the future implementation tasks. The detailed survey questions are available through https://docs.google.com/forms/d/e/1FAIpQLSdvDl-x-qR6MbLHrvtGKsomuDMQprmhQo6PVkMJnhSwu_zC-A/viewform?usp=sf_link

To promote the findings of this project and to demonstrate the potential of continuing its implementation, the research team has produced a downloadable software package. This package includes a demo video and a working copy of the project system that can be installed on any Windows 10 PC. It combines all the elements of the project code, facility detection models, and graphical user interface into a single, easy-to-install package. The tutorial in the demo video was included to demonstrate the pedestrian facility detection functions of this system on some example data. With this information, anyone with an interest in supporting the future implementation of this project is provided with a clear view of the current capabilities of the system and the ability to test it with their own data. This package can be downloaded here https://smttt-my.sharepoint.com/:f:/g/personal/w997046_usm_edu/EkEXWxihmUtHvejZteJDLZcBVhmLPLclGiJORDVFYXOuUg?e=dgS3Mh

During the development of this project, a next-step implementation was planned by the research team. This plan was discussed with the IDEA supervisor and an implementation plan proposal was formed in great detail. In this way, a clear path for improving the capabilities of the current system as demonstrated in the demo package was created. A potential grant opportunity to fund this implementation effort is NCHRP 20-44 Implementation Support Program, which has funding to facilitate implementation of completed NCHRP research results in state agencies. This project has proved the concept methods and developed the core functional models necessary for using machine learning and computer vision technology in multiple pedestrian facility detection and mensuration with high accuracy (shown in the blue area in Figure 16). However, implementing a system with these project results that will allow Departments of Transportation (DOTs) to collect ready-to-use information for sidewalks and crosswalks requires additional work related to both methodology and application (shown in the green area in Figure 16).

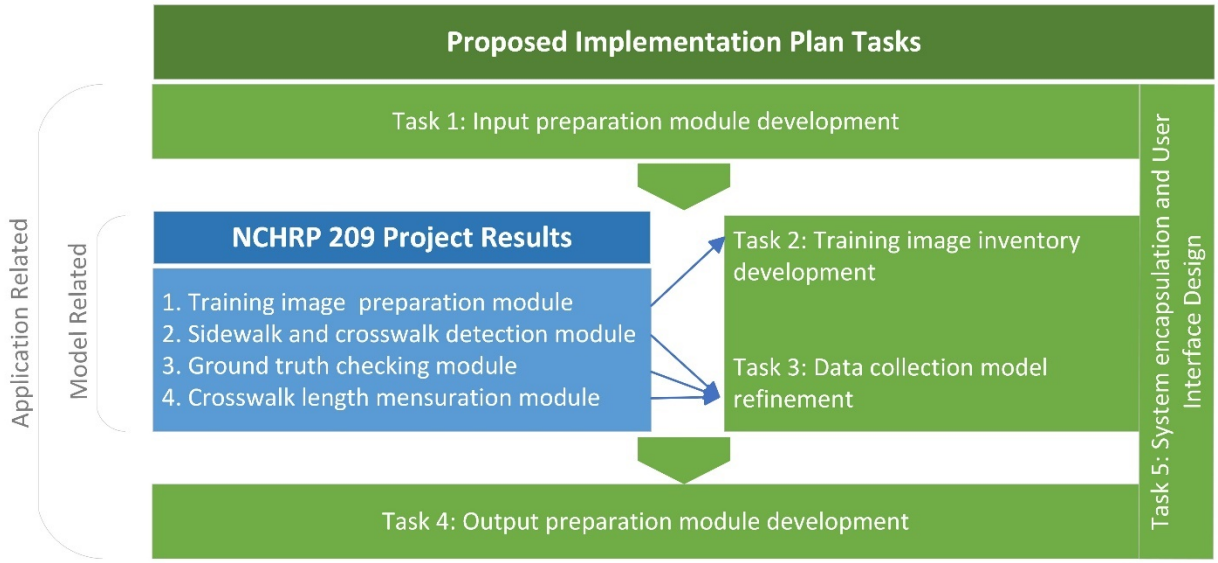


Figure 16: Proposed Implementation Plan Tasks

Presented below are five proposed tasks based on the discussion with the IDEA supervisor about the implementation plan.

Task 1. Design system input pre-processing module

The existing aerial images owned by DOTs cover large areas which contain many pedestrian facilities in different locations. To process these large area images, a workflow should be developed to pre-process the images so that only the aerial images of the candidate locations will be extracted and used as input for the data collection system where further processing will

be performed. The candidate locations would include locations where a crosswalk or a sidewalk could possibly exist. A few possibilities for this include 1) the location of every approaching lane at an intersection could possibly contain a crosswalk, 2) one side of a roadway segment could possibly contain a sidewalk, or 3) the roadway segment between two intersections could possibly contain a mid-block crosswalk. Narrowing down the input images to only focus on images of these candidate locations will reduce the processing time of the data collection system and decrease false detections by filtering out irrelevant but disruptive factors such as parking lots with parallel white lines.

Another challenging problem to solve in this task is to extract the aerial image surrounding a candidate crosswalk and ensure that the entire crosswalk would be contained in the image. Only in this way, can the length of the full crosswalk be measured automatically. The research team has already developed a "sliding window" method to adjust the extraction window automatically to cover the entire candidate crosswalk in one image. This method is necessary since deep learning models are only able to process images of a much smaller size than large area satellite imagery (256x256 pixel tiles in the tests here). Therefore, the imagery around candidate locations needs to be sliced into tiles intelligently.

This sliding window method works by detecting crosswalks that may be partially obscured by the boundary of the sliced image tile. Once a predicted bounding box is found near the edge of an image, the coordinates for the centerpoint of the partial detection are translated to the center of the current image before slicing a new image with gdal_translate. This has the effect of roughly centering the crosswalk candidate in the center of the image and allows the image to be reprocessed by the segmentation model. Figure 17 shows an example of this method being used to obtain a more accurate segmentation result for a crosswalk where the other curb was obscured by the tile boundary.

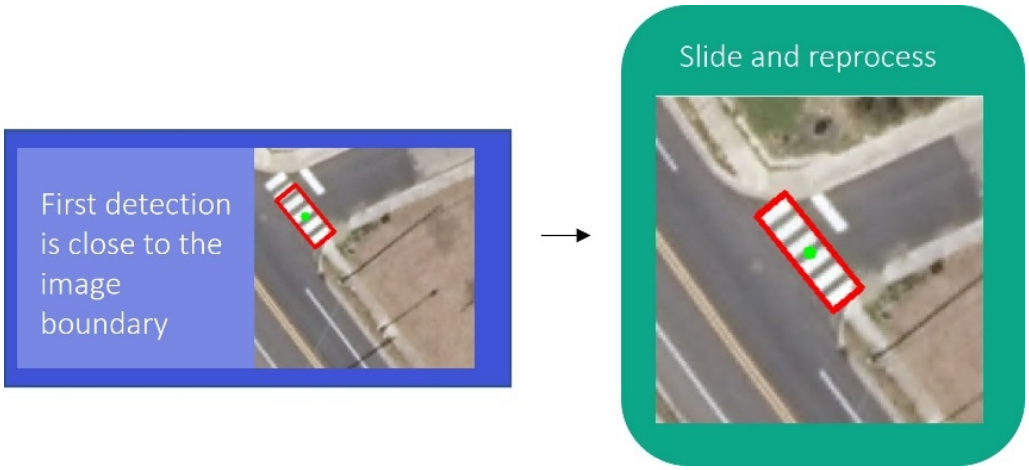


Figure 17: An example of the sliding window method automatically enhancing a segmentation result.

Task 2. Build a high-quality sample image repository.

This task will create a large image repository to prepare diverse and accurately tagged images for pedestrian facility detection. In the future, with the well-designed structure and standards of the repository that will be developed in this project, it could be easily expanded to include more types of facilities such as curb ramps, refuge islands, and push buttons on pedestrian signal lights, etc. Also, the availability of these additional facilities can conveniently expand the ability of the data collection system to collect more pedestrian facilities. Detailed tasks include: 1) Design the repository structure, standards, and format to not only store a large amount of image data but also to accommodate model training and easy expansion. 2) Enlarge the existing training sample image dataset collected by the NCHRP project. Utilize manual accuracy checking to guarantee accuracy. 3) Add more classifications of image sets. For example, “images with object in clear view”, “images of partial objects”, or “images with occlusion”. This task will greatly improve the capability of the detection models to handle special views of facilities in real-world images, such as facilities occluded by shadows from trees.

Task 3. Refine the detection models dedicated to sidewalks and crosswalks.

Although previous studies and existing applications have published many machine learning methods for object detection, few of them are dedicated to transportation facility detection. Directly using these existing models could cause the final system to suffer from inefficient and inaccurate pedestrian facility detection due to their reliance on overly general model architectures and unnecessarily complex parameters. In order to build a set of models

dedicated to transportation facility detection, novel and advanced deep learning-based models should be developed specifically based on transportation facility characteristics and unique features that can distinguish these facilities from other objects detected by popular detection methods.

Task 4. Determine and construct the system output sidewalk and crosswalk dataset.

Once the information for sidewalks and crosswalks is extracted, these data should be collected in a specific database and be ready to use in planning, designing, and managing. Thus, multiple tasks are required, including: 1) Determine the format of the output information. Options include shapefiles for sidewalks and crosswalks, excel tables for existing information of roadways with newly gathered sidewalk and crosswalk attributes, or a satellite table of sidewalks and crosswalks linked to the roadway inventories. 2) Design a mechanism to associate the newly collected sidewalk and crosswalk information with existing roadway and intersection information owned by the collaborating DOT. Sidewalk information on both sides of a roadway segment should be able to link to the same roadway segment. In addition, multiple crosswalks should be able to link to a single intersection or a roadway segment as mid-block crosswalks. 3) Design the schema diagram and tables in the database so that the information can be ready for use in pedestrian infrastructure planning, management, safety analysis, data query, etc. Also, information about newly built facilities added to the database in the future needs to be carefully considered in advance so as to maintain the data integrity in the database. 4) Build a function module to automatically generate the dataset in the format and structure determined and also to store the output information such as location of an identified sidewalk or crosswalk and the length of a crosswalk.

Task 5. Package and customize the computer application.

This task aims to package the above-mentioned models in a user-friendly form so that officers in the collaborating DOT can use it with limited training. Detailed steps include: 1) Determine the structure and platform of the application, considering computer software environment and software usage limitations required by the collaborating DOT. 2) Develop the application's platform, interface, and environment with the collaborating DOT for their internal use. 3) Pilot test the application with one district managed by the DOT, including at least one urbanized area and one rural area. 4) Deliver tools, brochures, and guides that can be used by the

DOT and other state DOTs for implementation. These materials will include step-by-step instructions on how to utilize the products with photos and documents.

5 CONCLUSIONS

The project developed a machine learning-based system to automatically detect, classify, and measure specific types of major pedestrian facilities, including sidewalks and crosswalks, from aerial images. The output of the system is information about the presence of sidewalks and crosswalks as well as crosswalk length and location for each processed image. To achieve this goal, the project has been implemented in two stages. Work in Stage I focused on developing a mechanism and functional models for automatically acquiring labeled aerial images, training the facility detection model using machine learning methods, improving the predictions for aerial images of occluded facilities by innovatively developing a dual-perspective model which uses aerial and street-view imagery simultaneously, and measuring the length of crosswalks. Work in Stage II completed the data collection system by integrating the functional models to evaluate the system's accuracy and efficiency by testing aerial images provided by MDOT. The following tasks completed in the project yielded exciting results:

1) A data acquisition workflow was developed to automatically prepare labeled sample data for sidewalk and crosswalk detection. This process automatically generated several large image datasets with images tagged as "having crosswalks", "not having crosswalks", "having sidewalks", and "not having sidewalks". The two main datasets used for development contain 4,380 images tagged for crosswalk detection and 29,450 images tagged for sidewalk detection. These labeled data served as the positive and negative samples that were used to train and test the facility detection model.

2) Based on these sample data, a prototype of each facility detection model has been developed using machine learning techniques. Specifically, a convolutional neural network (CNN) model was used to automatically detect and classify images into one of four classes (crosswalk, no-crosswalk, sidewalk, or no-sidewalk). These models were tested and achieved an accuracy rate of 96.35% ~ 98.43% for crosswalk detection and 92.87% for sidewalk detection, respectively. These testing results demonstrate the high accuracy and efficiency of collecting the data automatically with zero cost (not including the cost for tool development), compared to the cost of Caltrans' recent effort to award an Asset Collection Service Contract for millions of dollars.

3) Innovatively, to overcome situations where sidewalks or crosswalks are occluded in the aerial imagery, a dual-perspective mechanism was developed to double check the ground truth information for target objects by making use of both aerial and street-view images simultaneously. As a result, the ground truth checking model is now functioning for both sidewalks and crosswalks. A test on an image dataset with heavily occluded aerial crosswalk imagery showed that this model can increase detection accuracy by 49%.

4) The crosswalk mensuration model was developed using a dataset of 100 images that were manually prepared by the research team. This model can automatically obtain measurements such as crosswalk length by identifying a bounding box that contains all the pixels that belong to a crosswalk. In addition, the coordinates of the center of the bounding box are obtained and recorded as the location of the detected crosswalk.

5) A graphical user interface was created to encapsulate all models developed in the project to let users test images without knowing Python or any machine learning methodology. A test of 233 images from Forrest County Mississippi presented the accuracy of the system as high as 99.23% for crosswalk detection, 91.26% for sidewalk detection, and 93.7% for crosswalk length mensuration.

The project has proved the feasibility of using machine learning methods and image processing techniques to automate the data collection process for pedestrian facilities such as sidewalks and crosswalks. Detection methods and core functional models were developed based on deep learning and computer vision technology for performing multiple pedestrian facility detection and mensuration with high accuracy. Particularly, the application of deep learning methodologies, such as training a Convolutional Neural Network (CNN) to automatically detect crosswalks and sidewalks from images, showcased that deep learning-based methods enable knowledge extraction from images without requiring humans to manually select features beforehand. This also demonstrated that the deep learning method is more appropriate for handling the real-world conditions under which candidate images could contain clutter, shadows, saturation effects, distortion, occlusion, and many other unknown features. This makes the deep learning method surpass the previously used methods based on traditional image processing or machine learning methods.

Another contribution of this project to the field is that it provided an effective solution to solve the “occlusion” problem in real-world aerial images. Occlusion has been recognized as the

most challenging problem (*15*, *17*), causing omission of a crosswalk (*14*) during analysis or even malfunction of the algorithm. Occlusion of a crosswalk in an image could be caused by cars, trees, pedestrians, etc., as shown in . The suggested treatment in previous studies for the occlusion problem is to exclude images with occlusion during model development, which was common in global segmentation-based image processing approaches. As a result, the detection could only reach high accuracy when analyzing near-ideal pictures of crosswalks or sidewalks. Therefore, it was not applicable for generating data for an inventory of pedestrian facilities (*10*, *12*, *17*). In this project, to detect crosswalks and sidewalks from real-world images (especially with heavy occlusion), a dual-perspective, deep learning-based prediction method was developed to utilize the aerial view and street view of the same location simultaneously. Using this, occluded crosswalks can be verified automatically by checking the ground truth in their street view images using a combined model that takes advantage of both aerial images and street view images. One model was used as the initial detector for processing aerial images while a second one was used as an additional check of an alternative perspective (street-level) for the purpose of verifying the prediction made by the first detector. Combined, this ensemble model was proved to increase accuracy for occluded view detection by 50% (from 55.59% to 83.02%). More importantly, the recall value increased by 382.9% (from 15.41% to 74.42%), which means that a large majority of the occluded or unrecognizable crosswalks in the aerial view imagery were initially missed but were able to be recovered and correctly classified by the new method. However, this dual perspective method requires the availability of street view images at target locations. State DOTs need to provide either street view images or video logs to capture the images needed. Another solution to obtain street view level images is to query Google or Bing Maps through paid API calls.

The function models and system developed by this project form the foundation for developing a next-generation data collection method which could automatically detect, measure, and generate in-depth pedestrian facility information from images of the built environment on a large scale. The methods developed in this project not only promotes the application of automated methods for pedestrian facility data collection, but also provides a potential solution for detecting other transportation facilities, such as sidewalks or curb ramps, that are frequently occluded in real-world aerial images. The innovative ways of automating the data collection process provide “building blocks” for practitioners and researchers to adapt in building next-

generation data collection tools for automatically collecting specific infrastructure information of interest.

The implementation of the findings of this project faces several challenges that are worth being noted as lessons learned. The first challenge is the availability of the street view images of the candidate location. If the system is adopted by state DOTs, street view images should be provided by the DOTs internally, instead of using the street view images obtained from commercial databases like BingMaps or Google Maps. Even though some DOTs are able to collaborate with commercial image databases to extract their street view images, street view images are usually not available in many areas such as undeveloped areas, local roads, or gated subdivisions. One possible solution is to bring in crowdsourced street view images taken by pedestrians as a supplementary data source to cover areas without publicly available street view images. In addition, even when the street view images are available, in some cases the street view of a crosswalk or sidewalk can possibly be occluded by cars and other objects. Therefore, it is important to incorporate multiple angles of observation when designing a robust prediction system. Another problem with street view images is that their time stamp is not always the same as the corresponding aerial view images. There could be cases when the street view image was taken after a crosswalk was removed, but it might still be present when the aerial view image was taken (or vice versa).

The second challenge is regarding the quality of the training dataset of images tagged as having a crosswalk/sidewalk or not from OSM. The method developed here for extracting sample images of crosswalks and sidewalks from the large amount of aerial imagery covering a given area requires the crowdsourced tags for each crosswalk or sidewalk to accurately mark the location of the crosswalk or sidewalk. If the coordinates of a tag are not sufficiently close to the physical location of the crosswalk, the cropped images used as input for the facility detection model may not depict the actual target crosswalk or sidewalk. Also, nearby crosswalks or sidewalks can show up in an image that is marked as not having a crosswalk or sidewalk, which creates false positive samples. All these issues can lead to mismatched, incorrect images in both the training and testing data. In addition to the possibility of incorrect coordinates in crowd sourced data, it is also possible to obtain images that have a completely incorrect label. To investigate this, two evaluators looked through a subset of 1000 labelled crosswalk locations and 1000 labelled sidewalk locations from OSM and found the crowd sourced tags to be 78%

accurate on average for crosswalk labels and 94% for sidewalk. This preliminary result implies that there is a certain amount of incorrectly labeled images that likely affected the system performance adversely. Therefore, a training image repository would be recommended to be developed for the benefit of the automated data collection field. This repository, consisting of a large amount of diverse sample images of different types of pedestrian infrastructures accurately tagged by human evaluators with types and features, is critical to training a high-performance facility detection model. It is also recommended that additional, fine-grained classifications of image sets (e.g., "images with object in clear view", "images of partial objects", or "images with occlusion") should be included in the repository as well. This will greatly improve the capability of the detection models to handle special views of facilities in real-world images. Sample images in both aerial view and street view from imagery data available online or in government agencies should be obtained and manually tagged to guarantee high accuracy.

Except for the above-mentioned methodology related challenges, there are also certain application related challenges required for fully implementing the existing results. First, an input preprocessing module is required to extract aerial images from candidate locations to feed the data collection system. This is because that the aerial images owned by DOTs are of the entire district area instead of each target facility and pre-processing of these large area images needs to be conducted before feeding them directly into the models developed by the project. This task may include dividing an area image into small images containing just one approaching leg of an intersection or a short segment of a roadway. Second, as the output of the data collection system, information about sidewalks and crosswalks needs to be organized and stored in a ready-to-use format and structure by an output function module. Without any standard database structure for pedestrian facility data, a few questions need to be answered by a given DOT collaborator. For example, should the output be stored in a GIS layer or a table format? More importantly, how should the output be associated with the existing inventory of roadways and intersections owned by the DOT? Finally, the customized models, input preprocessing module, and output function module all need to be packaged into a user-friendly computer application so that the DOT officers can easily use the system without complex training.

By identifying remaining challenges toward full implementation, a clear path was formed through addressing each challenge as a task. The research team discussed the path and tasks with the IDEA supervisor and developed a detailed implementation plan as presented in Section 4 of

this document. The implementation plan will utilize the results from this project as the foundation for conducting another five tasks, including pre-processing district aerial image to generate images for detection, developing a training image inventory, refining data collection models, designing an output storage database, and the final task of encapsulating all components of the system and designing the user interface. Possible funding opportunities include NCHRP 20-44 Implementation Support Program, NCHRP Research Funding, the Research and Deployment Advisory Committee (RDAC) Funding, or Active Transportation Program (ATP) Funding.

In summary, this project has proved the feasibility and effectiveness of using deep learning methods and image processing techniques for automating pedestrian facility data collection from aerial images owned by DOTs. For any given aerial image of a candidate facility, the system can detect, classify, and measure the facility of interest, generating data about sidewalk presence, crosswalk presence, and crosswalk length obtained in that image. Based on the results of this project, an automated data collection tool can be developed for state DOTs to create an inventory of crosswalks and sidewalks in their jurisdictions, through solving the abovementioned challenges appropriately. It is worthy to note that this system and tool can be easily expanded to automatically collect data about much more pedestrian facilities, such as curb ramps, sidewalk gaps, and pedestrian activated signals, etc. Specifically, when trained appropriately, the methods can be used to collect facility condition such as pavement condition and striping condition. This would supplement the Active Transportation Asset Inventory Pilot (ATAIP) which is to collect bicycle and pedestrian asset information and facility conditions on the State Highway System. Additionally, data from video logs and crowdsourced data from sources other than OpenStreetMap could be incorporated to aid in future data collection efforts for this project. This would make possible the next generation of data collection tool that is fully automated and takes full advantage of the rich information buried in aerial images.

REFERENCES

1. Twaddell, H., E. Rose, J. Broach, J. Dill, K. Clifton, C. Lust, K. Voros, H. Louch, and E. David. *Fhwa Guidebook for Measuring Network Connectivity*. 2018.
2. Louch, H., K. Voros, and E. David. *Availability and Use of Pedestrian Infrastructure Data to Support Active Transportation Planning*. 2020.
3. Cevallos, F. *Safe and Accessible Pedestrian Facilities Inventory Model (SAPFIM): Development*. 2018.

4. Quiroga, C., and S. Turner. *ADA Compliance at Transportation Agencies : A Review of Practices*. 2008.

5. Zhang, Y., F. R. Proulx, D. R. Ragland, R. J. Schneider, and O. Grembek. *Develop a Plan to Collect Pedestrian Infrastructure and Volume Data for Future Incorporation into Caltrans Accident Surveillance and Analysis System Database*. Berkeley, CA, 2014.

6. Ahmetovic, D., J. M. Coughlan, R. Manduchi, and S. Mascetti. Zebra Crossing Spotter: Automatic Population of Spatial Databases for Increased Safety of Blind Travelers. *ASSETS 2015 - Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, 2015, pp. 251–258. https://doi.org/10.1145/2700648.2809847.

7. Senlet, T., and A. Elgammal. Segmentation of Occluded Sidewalks in Satellite Images. *Proceedings - International Conference on Pattern Recognition*, 2012, pp. 805–808.

8. Smith, V., J. Malik, and D. Culler. Classification of Sidewalks in Street View Images. *2013 International Green Computing Conference Proceedings, IGCC 2013*, No. June, 2013. https://doi.org/10.1109/IGCC.2013.6604476.

9. Hara, K., J. Sun, J. Chazan, D. Jacobs, and J. E. Froehlich. An Initial Study of Automatic Curb Ramp Detection with Crowdsourced Verification Using Google Street View Images. *AAAI Workshop - Technical Report*, Vol. WS-13-18, 2013, pp. 32–33.

10. Berriel, R. F., F. S. Rossi, A. F. de Souza, and T. Oliveira-Santos. Automatic Large-Scale Data Acquisition via Crowdsourcing for Crosswalk Classification: A Deep Learning Approach. *Computers and Graphics (Pergamon)*, Vol. 68, 2017, pp. 32–42. https://doi.org/10.1016/j.cag.2017.08.004.

11. Mattyus, G., S. Wang, S. Fidler, and R. Urtasun. HD Maps: Fine-Grained Road Segmentation by Parsing Ground and Aerial Images. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2016-December, 2016, pp. 3611–3619. https://doi.org/10.1109/CVPR.2016.393.

12. Riveiro, B., H. González-Jorge, J. Martínez-Sánchez, L. Díaz-Vilariño, and P. Arias. Automatic Detection of Zebra Crossings from Mobile LiDAR Data. *Optics and Laser Technology*, Vol. 70, 2015, pp. 63–70. https://doi.org/10.1016/j.optlastec.2015.01.011.

13. Luo, J., G. Wu. Developing an Interactive Machine-Learning-based Approach for Sidewalk Digitalization. *Project Report. Sponsored by California Department of Transportation*..

14. Wang, S., H. Pan, C. Zhang, and Y. Tian. RGB-D Image-Based Detection of Stairs, Pedestrian Crosswalks and Traffic Signs. *Journal of Visual Communication and Image Representation*, Vol. 25, No. 2, 2014, pp. 263–272. https://doi.org/10.1016/j.jvcir.2013.11.005.

15. Poggi, M., L. Nanni, and S. Mattoccia. Crosswalk Recognition through Point-Cloud Processing and Deep-Learning Suited to a Wearable Mobility Aid for the Visually Impaired. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9281, 2015, pp. 282–289. https://doi.org/10.1007/978-3-319-23222-5_35.

16. Mascetti, S., D. Ahmetovic, A. Gerino, and C. Bernareggi. ZebraRecognizer: Pedestrian Crossing Recognition for People with Visual Impairment or Blindness. *Pattern Recognition*, Vol. 60, 2016, pp. 405–419. https://doi.org/10.1016/j.patcog.2016.05.002.

17. Berriel, R. F., A. T. Lopes, A. F. De Souza, and T. Oliveira-Santos. Deep Learning-Based Large-Scale Automatic Satellite Crosswalk Classification. *IEEE Geoscience and Remote Sensing Letters*, Vol. 14, No. 9, 2017, pp. 1513–1517. https://doi.org/10.1109/LGRS.2017.2719863.

APPENDIX: RESEARCH RESULTS

NCHRP IDEA Program Committee

October 2021

Project Title: An Automated System for Pedestrian Facility Data Collection from Aerial Images

Project Number: NCHRP 209

Start Date: November 7, 2018

Completion Date: February 5, 2022

Product Category: New pedestrian facilities data collection methods

Principal Investigator:

Yuanyuan Zhang, Assistant Professor

Yuanyuan.zhang@usm.edu

Phone: 510-457-5689

Automate Pedestrian Facility Data Collection

Utilize deep learning and image processing techniques to automatically collect crosswalk and sidewalk data from aerial images

---

WHAT WAS THE NEED?

Although pedestrian facility data can improve pedestrian safety, it is not widely available at state level. Challenges, such as high cost for time and labor inherent in current data collection methods greatly impede the progress of collecting pedestrian facility data at a state level.

WHAT WAS OUR GOAL?

This project developed an innovative system to apply image processing and deep learning methods to automatically collect major pedestrian facility data, including sidewalk presence, crosswalk presence, and crosswalk length, from aerial images.

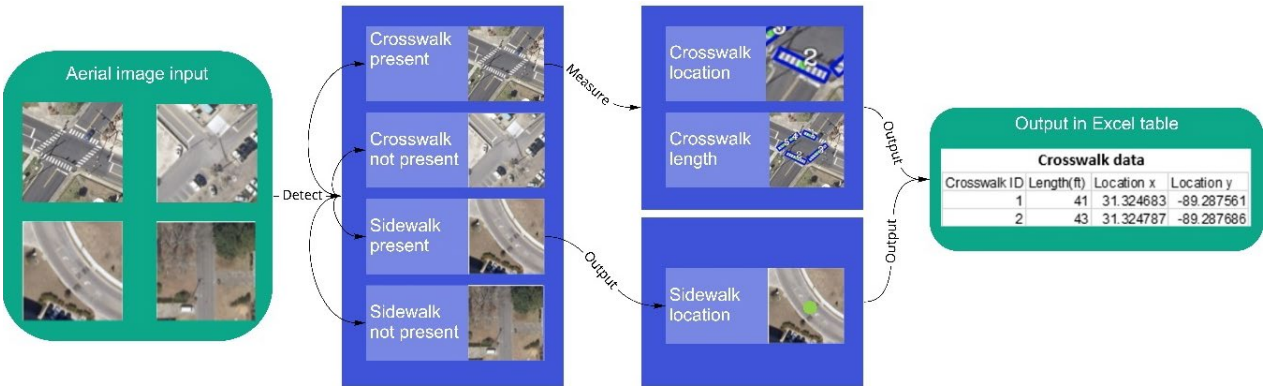
WHAT DID WE DO?

Using OpenStreetMap pedestrian facility location data and images from Bing Maps, four core models integrated in the system were developed (using convolutional neural networks based on

the VGG16 architecture) to prepare model development data (50,112 images), detect presence of facilities (with an accuracy of 96.35% ~ 98.43% for aerial crosswalk images, and 92.87% for aerial sidewalk images), check the ground truth when street view data are available, and measure the length of the detected crosswalk. The aerial image provided by the Mississippi Department of Transportation was tested to evaluate the validity of the system on unseen data.

WHAT WAS THE OUTCOME?

A pioneering system (as shown in below) was developed to process aerial images of candidate locations. Final outputs are then stored in an excel table. In the test, 400 images from Forrest County, Mississippi were processed, resulting in an accuracy as high as 99.23% for crosswalk detection, 91.26% for sidewalk detection, and 93.7% for crosswalk length mensuration.



The product - automated data collection system, in a picture

WHAT IS THE BENEFIT?

Using the system developed in this project, the time cost to collect data from an area of 466.31 square miles would be approximately 8 days on a GPU workstation pc. This automated data collection method has potential to greatly decrease the monetary and time cost of state-level pedestrian facility data collection and form the next generation data collection method.

LEARN MORE

The report is available through IDEA program website. The project code, user interface, and a video tutorial are accessible through https://smttt-my.sharepoint.com/:f:/g/personal/w997046_usm_edu/EkEXWxihmUtHvejZteJDLZcBVhmLPLclGiJORDVFYXOuUg?e=dgS3Mh