THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

An AI-enabled Annotation Platform for Storefront Accessibility and Localization

Xuan Wang[3], Jiawei Liu[1], Hao Tang[2,3], Zhigang Zhu[1,3], William H. Seiple[4]
[1]Department of Computer Science, The CUNY City College
[2]Department of Computer Information Systems, The CUNY BMCC
[3]Department of Computer Science, The CUNY Graduate Center
[4]Lighthouse Guild, New York, NY 10023
xwang4@gradcenter.cuny.edu, jliu021@citymail.cuny.edu,
htang@bmcc.cuny.edu, zzhu@ccny.cuny.edu, wseiple@lighthouseguild.org

Abstract

Although various navigation apps are available, people who are blind or have low vision (PVIB) still face challenges to locate store entrances due to missing geospatial information in existing map services. Previously, we have developed a crowdsourcing platform to collect storefront accessibility and localization data to address the above challenges. In this paper, we have significantly improved the efficiency of data collection and user engagement in our new AI-enabled *Smart DoorFront* platform by designing and developing multiple important features, including a gamified credit ranking system, a volunteer contribution estimator, an AI-based pre-labeling function, and an image gallery feature. For achieving these, we integrate a specially designed deep learning model called *MultiCLU* into the Smart DoorFront. We also introduce an online machine learning mechanism to iteratively train the MultiCLU model, by using newly labeled storefront accessibility objects and their locations in images. Our new DoorFront platform not only significantly improves the efficiency of storefront accessibility data collection, but optimizes user experience. We have conducted interviews with six adults who are blind to better understand their daily travel challenges and their feedback indicated that the storefront accessibility data collected via the DoorFront platform would be very beneficial for them.

Keywords

Crowdsourcing, Storefront Accessibility, Independent Travel, Visually Impaired, Open-Source Data, Deep Learning.

Introduction

As reported by multiple sources (CDC, 2021; IAPB, 2022), over four million people living in the United States are blind or have low vision, and the prevalence of vision loss is predicted to increase dramatically in the next decades. The limitations caused by vision loss in daily life are unimaginable, even travel can be considered as the one of the most stressful events in the lives of people with visual impairment or blindness (PVIB) (Donaldson, 2017). Difficulties with travel limit essential activities, such as visiting local stores and using public transportation facilities, etc. In addition to the obstacles presented by many transportation options, proximal information about storefronts is essential for access and poorly designed entrances place PVIB at risk (MOPD, 2022; ADA, 2022). Therefore, knowing the accessibility data of the storefronts in advance can significantly alleviate their fear in exploring their local community.

In order to gather storefront accessibility and localization image data in a complicated street level environment, to support independent travel of PVIB, we have developed *DoorFront* (Liu, et al., 2022), a crowdsourcing web platform that collects large-scale storefront accessibility and localization data in New York City. Our web-based platform DoorFront allows volunteers to remotely label storefront accessibility objects on Google Street View panoramas (Google, 2022) in four main categories: *door, doorknob, stair* and *ramps*. The feedback from volunteers have demonstrated its usability and high potential for data collection.

However, the data collection of large-scale storefront accessibility and localization data using crowdsourcing is still very labor intensive. Therefore, in this paper, we introduce an AI-enabled annotation platform for storefront accessibility and localization, which is built on our previous DoorFront platform. In order to improve user experience, user label efficiency and quality, we implemented a number of new features into the DoorFront platform, including gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. As a key component, we integrate a specially designed deep learning model called *MultiCLU* (Wang, et al., 2022), for enabling label automation. Furthermore, we introduced an iterative training mechanism for our MultiCLU model to provide better labeling performance and 'autowalker,' an enhanced pre-label process to further improve the efficiency of human labeling. To summarize, the contributions in this work include:

1. An AI-enabled annotation platform for storefront accessibility and localization, *Smart DoorFront*, which integrates new features, including credit ranking, volunteer contribution estimating, AI-based pre-labeling, and image gallery.
2. Enabling label automation by integrating a specifically designed deep learning model - MultiCLU.
3. An iterative online training mechanism for our deep learning model to provide better labeling performance and user experience.

Related Work

Crowdsourcing Accessibility Data

Accessibility data, including sidewalk accessibility data, storefront accessibility data, and public infrastructure data, is essential information for PVIB in order to plan independent travel. To collect up-to-date accessibility data, local and state governments, and even federal government, often conduct street audits to inquire about specific conditions (May et al., 2014; Law et al., 2018). However, manual audits are both time consuming and expensive. One of the novel alternatives is the use of crowdsourcing techniques. Many recent studies have demonstrated the feasibility of crowdsourcing approaches (Krajzewicz et al., 2010; Marzano et al., 2019), in collecting useful information on urban mobility and public transportation. Hara and his team have proposed methods combining Google Street View and online crowdsourcing to provide sidewalk accessibility data and bus stop locations (Hara et al., 2013, 2014). However, there are few efforts focused on storefront accessibility and localization information.

Object Detection in Urban Environment

Many approaches have been proposed for object detection in urban scenes. Some of them (Du et al. 2012; Sabir et al. 2018; Zhu et al. 2016) are for text and signage detection in street level environments. Ahmetovic, et al. (2015) mined existing spatial image datasets for discovery of zebra crosswalks in urban environments, which could increase safety of PVIB when traveling. Sun et al. (2017) used deep learning to find missing curb ramps at city street regions, which could help not only PVIB, but also people with mobility disabilities. Weld, et al. (2019) propose a deep learning framework to automatically detect sidewalk accessibility using streetscape imagery. To our best knowledge, there are few, if any, studies related to detect storefront accessibility and localize them using street level imagery. In this paper, we integrate our specifically designed deep learning model, MultiCLU (Wang et al. 2022), to enable pre-label automation for our DoorFront platform by using our previously collected storefront accessibility and localization image dataset.

Discussion

User Study

In order to ensure the effectiveness and usefulness of our Smart DoorFront platform, we conducted informal interviews with six adults who are blind to better understand their challenges when they accessed essential activities. Table 1 shows the doorfront interview questions and the answers from the interviewees. We started with basic travel questions (Q1 to Q4 in Table 1), followed by storefront accessibility related questions (Q5 to Q11 in Table 1).

Table 1. Doorfront Interview Questions and Answers.

Questions	**Answers**
1.How often do you travel?	Every day: 100% Once a week: 0% Once a month: 0%
2.What is your major transportation while you are traveling?	Access-a-Ride: 17% Subway: 83% Walking: 0%
3.How far do you often travel?	<1 mile: 0% 1 - 5 miles: 83% > 5 miles: 17%
4. How long do you spend per trip?	< 1 hour: 17% 1 - 3 hours: 83% 3 - 5 hours: 0% > 5 hours: 0%
6.Which store do you often visit (Multiple choices)?	Grocery store: 100% Supermarket: 67% Restaurant: 0% Shopping mall: 0%
7.How difficult is it to find the entrance of the store (Rate from 1 to 5)?	1 (very easy): 0% 2 (easy): 0% 3 (normal): 0% 4 (hard): 83% 5 (very hard): 17%
8.What is the most challenging task when you arrive at the entrance of a store?	Locate the entrance; Door type for the entrance. Time consuming if no one helps.
9.What kind of assistance do you expect when you arrive at the entrance of a store?	Direction guidance: 17% Lead to the entrance: 83%
10.Our current application can locate and detect three main storefront objects: door, knob and stair, with estimated location of knob, do you think they are sufficient?	Yes: 100% No: 0%
11.What do you expect to know when you arrive at the entrance of a store?	Accurate location for the door and stair. Relative location for the knob.

Through this series of interviews, we learned that most of the participants were daily commuters, and that they travel independently. All participants often visited grocery stores or supermarkets, but none of them visited restaurants and shopping malls (Q6 in Table 1). They stated that finding the locations of store entrances is very challenging for them (Q7 in Table 1). Two participants mentioned their previous experiences of finding a store entrance. They specifically indicated that one of the most challenging tasks is to find the entrance with a glass door. There are many buildings with glass walls at the ground floor in NYC, our participants indicated that it is very challenging to find the glass door that hides inside the glass walls. If no one provides help, they could spend a great amount of time finding and locating the entrance. During the interview, we found that each participant needs different levels of assistance to find the entrance of a store, but most of them prefer a person who can lead them to the entrance directly, instead of providing verbal guidance or direction. Most participants indicated that knowing the location of the entrance (e.g., a door) and the access of the entrance (e.g., the stairs to the door) were very important for them, which could significantly reduce their travel time and relieve their stress. They also agreed that providing a relative location for the doorknob to the door (e.g., "*The knob is on the left side of the door.*") could work for them if the location for the entrance is accurate.

The feedback from the interviews indicated that helping PVIB to identify the accessibility of the storefront would be very beneficial for them, and could greatly ease their daily burdens and improve their independence. Not only we received valuable feedback for better understanding challenges for PVIB through the survey, but we also obtained ideas on expanding the categories of storefront accessibility objects (e.g., handrails for the stairs) and improving our platform.

Doorfront Platform Improvements: An Overview

DoorFront is a web-based application that combines Google Street View and crowdsourcing, with an interactive interface and a user-friendly labeling tool (Liu, et al., 2022). There are two main pages in the original DoorFront, namely an *Exploration* page and a *Labeling* page. Volunteers not only can virtually walk-through New York City with embedded interactive Google Street View provided on the Exploration page; but they can also label storefront accessibility data with the functional and user-friendly labeling tool. Even though the feedback from crowd volunteers has demonstrated its usability and high potential for data collection, the process of labeling is still relatively labor intensive. Our studies show that there are two key factors that influence the effectiveness of the data collection: the number of volunteers participating in DooFront and the time they spent in the labeling process.

To address these issues, we have made significant improvements to the DoorFront application, leading to *Smart DoorFront*, which includes four major new features: gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. We will discuss each of them in the following.

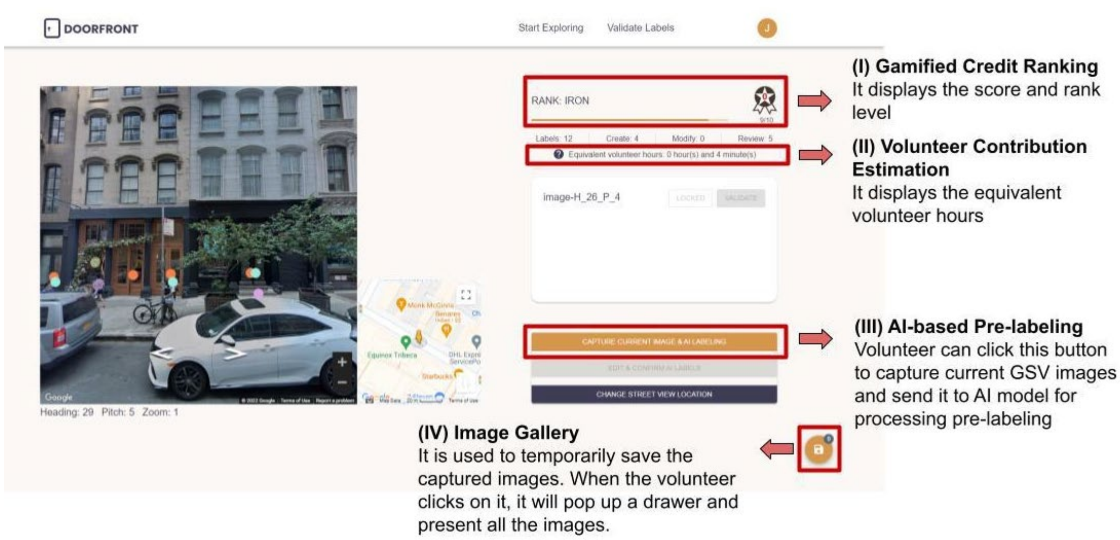

Fig. 1. The Interface of the *Smart DoorFront* Exploration Page.

Gamified Credit Ranking

Inspired by gamified settings (Ning, 2018), we designed and implemented a *gamified ranking system* (Fig 1, part I) and a leaderboard for the volunteers. In this project, we define seven different rank levels and their corresponding cumulative credits. The specific names of each level and the details of the accumulated credits for each phase are shown in Table 2.

Table 2. Rank levels and accumulated credits.

Level	Iron	Bronze	Silver	Gold	Platinum	Diamond	Challenger
Credits	0 - 9	10 - 49	50 - 299	300 - 999	1000 - 1999	2000 - 4999	5000 - 9999

Volunteers will receive credits through three contributed operations: annotating new Google Street View images, correcting other volunteers' annotations, and reviewing other volunteers' annotations. To further increase the entertaining nature of DoorFront, we also develop a treasure hunt feature that allows volunteers to earn extra ten credits whenever they find a treasure; once DoorFront is initialized, all treasures will automatically be hidden in random areas of New York City. Therefore, the locations of the treasures are completely different each time, and volunteers are unable to gain these extra points by memorizing the locations of the treasures. With respect to leaderboard, it will show volunteers according to their levels. In this atmosphere, we believe volunteers will become competitive and spend more time in data collection to advance their level.

Volunteer Contribution Estimation

Crowdsourcing brings us flexibility so we can distribute the data collection tasks to volunteers. Volunteers from each borough in New York City can collect storefront accessibility data in their own communities through DoorFront. However, using the original DoorFront platform, we were unable to recruit a large number of volunteers to participate. We needed to address how to attract more volunteers to DoorFront. In the initial version, we decided to award volunteer certificates to volunteers through our collaboration with Lighthouse Guild, a vision and health care organization. They provided volunteer certificates to individuals who have made large contributions of time on the platform. However, the algorithm we used did not work well to calculate the equivalent volunteer time. The core idea was calculating the number of images collected by the volunteers. We assume that, on average, volunteers spend one minute to annotate each image without the assistance of the AI model. The shortcoming of this algorithm is obvious. Since the number of labels in each image is different, our algorithm does not reflect well the effort of volunteers and the time they spend.

Therefore, we decided to design a new *volunteer contribution estimator* to better calculate volunteer effort. First, we rebuilt the DoorFront's credit management system. With this improvement, we are now able to monitor the number of labels annotated by volunteers, and equivalent-volunteer-time is determined by the number of labels. Second, we implemented a small widget to showcase the volunteer's effort in real-time (Fig. 1, part II). In addition, to further encourage volunteers to promote our application, we also provide sharing buttons on different social media applications such as Meta and Twitter, to share their contributions with their friends. With these improvements, we believe that more and more younger volunteers, especially middle or high school students, will be interested in participating in our study.

AI-based Pre-labeling

One of the key issues we needed to address is to reduce the time for annotation by a volunteer. On average, it takes at least one minute to manually annotate a storefront image from scratch using our DoorFront interface. There are three steps in the annotation process: (1) identify a storefront accessibility object; (2) annotate the object with a bounding box; and (3) add a subtype if the object is a door or door handle. Volunteers need to repeat these three steps until they label all the storefront accessibility objects in a scene, hence the labeling task is still time-consuming.

In order to further improve the efficiency of data collection, we enhanced *DoorFront* with an *AI-based pre-labeling* function (Fig. 1, part III). With AI support, DoorFront can perform pre-labeling once a volunteer captures a new Google Street View image. This means that they do not need to label the image from scratch and the only thing they need to do is validate the results predicted by our AI model. Compared to the initial workflow, we now have only two steps: (1) Verify and correct the annotations labeled by the AI model; and (2) Add subtypes. Based on the outstanding performance of our model, we can skip the first step in most cases, which dramatically reduces the annotation time.

In addition, we introduced an *autowalker*, a pre-label process to further improve the efficiency of human labeling. Our autowalker will automatically walk through the Google Street View in the direction selected by a volunteer, capture 50 images, and use our model to pre-label the storefront accessibility data. Meanwhile, our autowalker will also collect the GIS information (latitude, longitude) and corresponding heading, pitch and zoom level.

Image Gallery

To maximize the utilization of the AI model, we modified the way that we save Google Street View images. In the initial version, DoorFront only allowed volunteers to label one image at a time. Now with the Smart DoorFront, volunteers can capture multiple images while they are virtually walking along with the street. Those images will be temporarily stored in an *image gallery* (Fig 1, part IV) and then sent to the AI model for pre-labeling processing. Volunteers can then validate all images at once, without frequently switching among different web interfaces, which greatly reduces their labeling time.

Furthermore, we store the untagged images in our remote database. With this information, volunteers will be able to access these images again, regardless of the last time they exited the application. Furthermore, our application will send notifications to remind volunteers that they forgot to annotate these images.

In the next sections, we will describe the enabler of the aforementioned features: the integration of the deep learning model and an iterative learning approach.

Deep Learning Model Integration

In order to improve the efficiency of the storefront accessibility labeling process, we integrate a specially designed deep learning model MultiCLU (Wang, et al., 2022) into DoorFront. Our MultiCLU model is implemented by integrating the state-of-the-art object detector, Faster R-CNN (Ren, et al., 2015), with contextual relationships among storefront accessibility objects, in order to improve the accuracy of image detections (Fig. 1, part III). For example, if we know there is a knob in the image, we can easily guess there should be a door in the image. Furthermore, a knob must appear inside a door frame, and if a stair exists, it should be under the door. Our MultiCLU model utilizes these contextual relationships to improve our detection results. The overall pipeline of MultiCLU in DoorFront is shown in Fig. 2. When a volunteer captures an image, MultiCLU will detect and localize storefront accessibility objects within a few seconds before user labeling. Volunteers can then validate or edit the labels which are pre-labeled by MultiCLU. Our platform will record three main types of labels: (1) Added labels from volunteers (Fig. 2. part I); (2) Removed labels from volunteers (Fig. 2. part II); and

(3) Validated AI labels (Fig. 2. part III). Our model is further trained on modified labels which are corrected by volunteers, to further improve the performance.

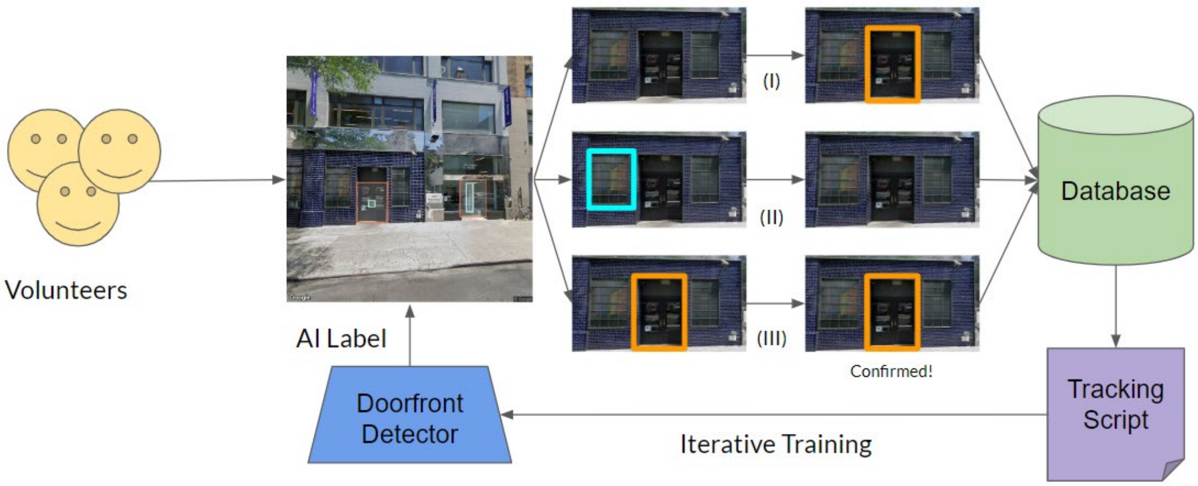

Fig. 2. Overall pipeline of MultiCLU-based labeling in DoorFront.

Dataset Description

We used 1232 storefront accessibility images previously collected by our DoorFront platform for training our initial MultiCLU model. Among them, there are 604 images which have labels from multiple volunteers. We used the average labels from these images as our base training set. The remaining images were used as a testing set. We did not use the ramp category for our training because of the lack of the labels. Table 3 and Table 4 show the summary of the dataset. The precision and recall for the three categories are shown in the "Initial Model" in Table 5 and Table 6, respectively.

Table 3. The training set of collected storefront accessibility dataset over time.

Class	Initial Training set	After Day 1	After Day 2	After Day 3
Door	1225	1532	1719	1913
Knob	863	962	1060	1173
Stair	270	346	422	475

Table 4. The testing set of collected storefront accessibility dataset over time.

Class	Initial Testing set	After Day 1	After Day 2	After Day 3
Door	1080	1132	1168	1229
Knob	887	905	928	979
Stair	197	209	224	240

Iterative Training

We also introduced a training automation mechanism to iteratively train our MultiCLU model. The model will start iterative training automatically when a certain amount (***N***) of new images has been recorded with new labels. As shown in Fig. 3, if we achieve better performance after training, we will replace the current model, otherwise we will start another training process the next day and keep the current model. We also use a data aggregation process to improve the robustness of the detection. When the $^{(n+1)}/[\ \ /_{?}]$ *th* iteration starts, we accumulate the previous training dataset from n [$/_{?}$] *th* training step, where n denotes current training step.

And then we use the combined dataset to refine the current model.

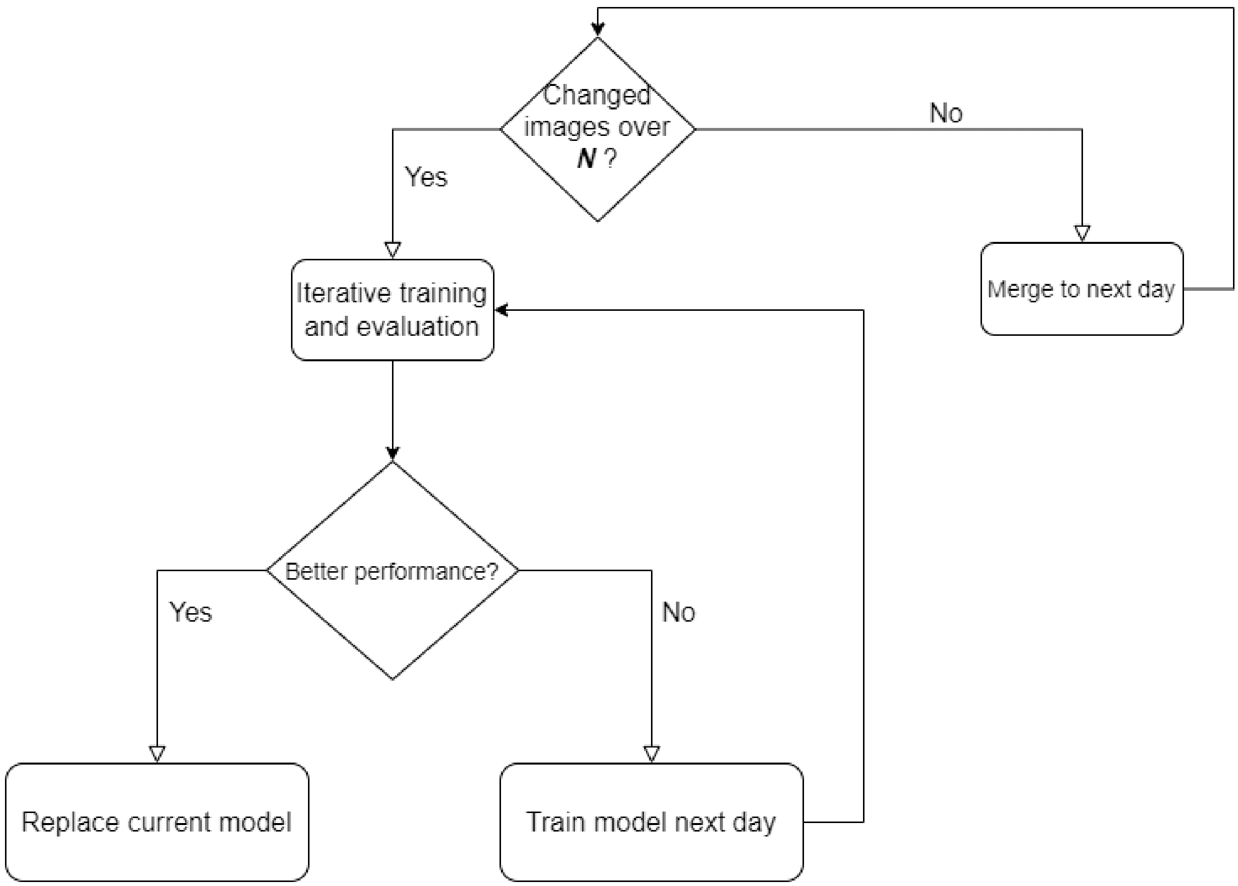

Fig. 3. Iterative training process with MultiCLU and DoorFront.

Model Evaluation

We evaluated our iterative training mechanism using labels collected within three consecutive days. For new labels from each day, we randomly select 80 percent of the dataset as the training set to refine the previous MultiCLU model, and the remaining 20 percent were added to the current testing set (Table 3 and Table 4). We accumulated both training and testing data into our previously collected storefront accessibility and localization dataset. We only used the accumulated labels from three consecutive days to refine our model, which could improve the robustness of the detection. We report both precision (Table 5) and recall (Table 6) of the 3-day iterative training. We observed that both precision and recall for all the categories were improved, with precision increasing from +1.1% to +3.7% and recall from +1.3% to +5.2%, respectively.

Table 5. The precision for the initial model and 3-day iterative trained models.

Class	Initial Model	Day 1	Day 2	Day 3
Door	78.7%	79.5%	83.2%	83.2%
Knob	79.0%	78.8%	81.6%	82.7%
Stair	81.2%	81.2%	81.6%	82.3%

Table 6. The recall for the initial model and 3-day iterative trained models.

Category	Initial Model	Day 1	Day 2	Day 3
Door	88.2%	88.9%	89.7%	90.2%
Knob	85.4%	85.4%	86.7%	88.0%
Stair	77.6%	78.0%	82.6%	83.2%

Conclusions

Building on our previous DoorFront platform, our new platform, Smart DoorFront, integrates gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. We utilize our specially designed storefront image detection model MultiCLU, which is built upon the state-of-the-art object detector and uses of context information among storefront accessibility objects. In addition, we introduce an autowalker, a pre-label process to further improve the efficiency of human labeling. We also introduce an iterative training mechanism to improve the accuracy and robustness of our deep learning model. Our new platform not only optimizes user experience, but also significantly improves the efficiency of storefront accessibility labeling process with our deep learning model. We will continue gathering feedback from volunteers and develop a mobile app for PVIB to navigate to store entrances, using the collected storefront accessibility and localization data. We will also integrate our deep learning model into a mobile app in the future, to better help PVIB to improve their independence. We will then perform a formal user evaluation of storefront accessibility and localization using the mobile app to understand how much the app could improve their performance, and what improvements we need for the algorithms and mobile app.

Acknowledgments

The research is supported by the National Science Foundation (NSF) through Awards #2131186 (CISE-MSI), #1827505 (PFI), and #1737533 (S&CC). The work is also supported by the US Air Force Office of Scientific Research (AFOSR) via Award #FA9550-21-1-0082 and the ODNI Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers University (#HHM402-19-1-0003 and #HHM402- 18-1-0007).

Works Cited

Ahmetovic, Dragan, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. "Zebra crossing spotter: Automatic population of spatial databases for increased safety of blind travelers." *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 2015.

The Americans with Disabilities Act Checklist for Readily Achievable Barrier Removal (ADA), https://www.ada.gov/racheck.pdf, last visited Aug 2022.

Cheng, Mingmei, Yigong Zhang, Yingna Su, Jose M. Alvarez, and Hui Kong. "Curb detection for road and sidewalk detection." *IEEE Transactions on Vehicular Technology* 67, no. 11 (2018): 10330-10342.

Donaldson, Nadine. "Visually Impaired Individuals' Perspectives on Obtaining and Maintaining Employment." (2017).

Du, Yuning, Genquan Duan, and Haizhou Ai. "Context-based text detection in natural scenes." *19th IEEE International Conference on Image Processing. IEEE*, 2012.

"Fast Facts of Common Eye Disorders." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 9 June 2020, www.cdc.gov/visionhealth/basics/ced/fastfacts.htm

"Global Estimates of Vision Loss." The International Agency for the Prevention of Blindness, 31 May 2022, https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/.

Google Map API (2022), https://developers.google.com/maps, last visited Aug 2022

Hara, Kotaro. "Scalable methods to collect and visualize sidewalk accessibility data for people with mobility impairments." *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology*. 2014.

Hara, Kotaro, Shiri Azenkot, Megan Campbell, Cynthia L. Bennett, Vicki Le, Sean Pannella, Robert Moore, Kelly Minckler, Rochelle H. Ng, and Jon E. Froehlich. "Improving Public Transit Accessibility for Blind Riders by Crowdsourcing Bus Stop Landmark Locations with Google Street View." *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 2013, https://doi.org/10.1145/2513383.2513448.

Krajzewicz, Daniel. "Traffic Simulation with SUMO - Simulation of Urban Mobility." *Fundamentals of Traffic Simulation*, 2010, pp. 269–293., https://doi.org/10.1007/978-1-4419-6142-6_7.

Law, Stephen, Brooks Paige, and Chris Russell. "Take a Look Around: Using Street View and Satellite Images to Estimate House Prices." *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 5, 2019, pp. 1–19., https://doi.org/10.1145/3342240.

Liu, Jiawei, Hao Tang, William Seiple, and Zhigang Zhu. "Annotating Storefront Accessibility Data Using Crowdsourcing." *Journal on Technology and Persons with Disabilities* Volume 10, California State University, Northridge., 1 Jan. 1970, https://scholarworks.csun.edu/handle/10211.3/223472.

Marzano, Gilberto, Joanna Lizut, and Luis Ocha Siguencia. "Crowdsourcing Solutions for Supporting Urban Mobility." *Procedia Computer Science*, vol. 149, 2019, pp. 542–547., https://doi.org/10.1016/j.procs.2019.01.174.

May, Andrew, Christopher J. Parker, Neil Taylor, and Tracy Ross. "Evaluating a Concept Design of a Crowd-Sourced 'Mashup' Providing Ease-of-Access Information for People with Limited Mobility." Transportation Research Part C: Emerging Technologies, vol. 49, 2014, pp. 103–113., doi:10.1016/j.trc.2014.10.007.

Ning, Bing. “Aux-driven design method for building gamification system.” International Conference of Design, User Experience, and Usability. Springer, Cham, 2018.

NYC Mayor's Office for People with Disabilities (MOPD) https://www1.nyc.gov/site/mopd/index.page, last visited Aug 2022

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. “Faster r-cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems 28 (2015).

Sabir, Ahmed, Francesc Moreno-Noguer, and Lluís Padró. “Enhancing text spotting with a language model and visual context information.” *In Proceedings of the Computer and Communications Industry Association*, 2018.

Sun, Jin, and David W. Jacobs. “Seeing what is not there: Learning context to determine where objects are missing.” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5716-5724. 2017.

Wang, Xuan, Jiajun Chen, Hao Tang, and Zhigang Zhu. “MultiCLU: Multi-stage Context Learning and Utilization for Storefront Accessibility Detection and Evaluation.” 2022 International Conference on Multimedia Retrieval, https://dl.acm.org/doi/pdf/10.1145/3512527.3531361

Weld, Galen, Esther Jang, Anthony Li, Aileen Zeng, Kurtis Heimerl, and Jon E. Froehlich. “Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery.” *In The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 196-209. 2019.

Zhu, Anna, Renwu Gao, and Seiichi Uchida. “Could scene context be beneficial for scene text detection?” *Pattern Recognition* 58 (2016): 204-215.