PathLookup: A Deep Learning-Based Framework to Assist Visually Impaired in Outdoor Wayfinding

Uddipan Das
Department of Computer Science
The College of New Jersey
Ewing, NJ 08628, USA
dasu@tcnj.edu

Vinod Namboodiri
Department of EECS
Wichita State University
Wichita, KS 67260, USA
vinod.namboodiri@wichita.edu

Hongsheng He
Department of EECS
Wichita State University
Wichita, KS 67260, USA
hongsheng.he@wichita.edu

***Abstract*—Reading and following visual signs remains the predominant mechanism for navigation and receiving wayfinding information in areas without accurate GPS coverage. This puts people who are blind or visually impaired (BVI) at a great disadvantage. There still remains a great need to provide a low-cost, easy to use, and reliable auxiliary wayfinding system within indoor and outdoor spaces that complements existing satellite-based systems. Through both a user study and a quantitative study of GPS accuracies in outdoor environments, this paper highlights the need for auxiliary outdoor wayfinding tools for people with visual impairments. A deep learning-based image localization framework called PathLookup is proposed in this work for accurately providing path advancement information for outdoor wayfinding. Evaluation results show PathLookup to be highly accurate and fast potentially proving to be a valuable tool for future integration into outdoor wayfinding systems.**

***Index Terms*—Accessibility technologies, outdoor wayfinding, deep learning, computer vision, visually impaired persons**

I. Introduction

Wayfinding is a critical task for independent living. Wayfinding remains a challenge for people with disabilities in our communities. For outdoor environments, recent advances in satellite-based technologies (e.g. GPS) along with the pervasiveness of smartphones provide an accurate and simple to use means for wayfinding. However, there remain many outdoor areas such as sidewalks, within and around office buildings, public recreational areas, and university campuses, where the effectiveness of satellite-based systems such as global positioning systems (GPS) is limited or non-existent. In areas with decent GPS coverage, often its accuracy is not good enough for fine-grained pedestrian navigation. Existing GPS-enabled navigation systems in mobile phones provide a median of 5.0 to 8.5 m error in localization [1]. Furthermore, wayfinding remains a challenge in many indoor environments, especially those that are geographically large, such as grocery stores, airports, sports stadiums, office buildings, and hotels. Reading and following visual signs remains the predominant mechanism for receiving wayfinding information in areas without or inaccurate GPS coverage. This puts people who are blind or visually impaired (BVI) at a great disadvantage. Thus, there still remains a great need to provide a low-cost, easy to use, and reliable wayfinding system within indoor and outdoor spaces that complements existing satellite-based systems.

Recent work in developing systems for wayfinding use low-cost, stamp-size BLE "beacon" devices embedded in the environment that interact wirelessly with smartphones carried by users are promising in indoor environments [2]–[5] but incur potentially large infrastructure costs to get adequate coverage outdoors. The use of computer vision techniques to provide critical inputs about the whereabouts of a person as they navigate in outdoor auxiliary wayfinding environments can reduce the reliance on infrastructure-based solutions (and their associated costs). Some recent work in this direction has appeared through the application of computer vision-based techniques [6], [7]. These, have however been limited efforts without solving some of the fundamental challenges in pedestrian outdoor navigation.

For outdoor navigation, a fundamental building block is to understand where a user is and which direction they should advance next from their current location. Given GPS inaccuracies, a major area of concern for pedestrian navigation is recognizing the current path and where it is leading. This is especially challenging at intersections where paths diverge to possibly different destinations. This paper presents a deep learning-based computer vision framework called PathLookup that allows a user to capture an image in the direction of the path they are taking (or intend to take) and receive details about it. PathLookup uses a convolutional neural network (CNN) to predict a path en-route to a destination and utilizes a pre-constructed contextual database to provide associated contextual information. This paper presents the design, implementation, and evaluation of PathLookup as a proof-of-concept in a university setting where pedestrian navigation is of utmost importance. Evaluation results show that the CNN network of PathLookup performs with a 99.5% accuracy on a validation dataset when identifying paths toward a destination.

II. Related Work

There have been some recent efforts in Bluetooth-based indoor localization, using BLE beacons for wayfinding for the BVI, such as StaNavi [2], GuideBeacon [5], ASSIST [8], PerCept [9], and NavCog [3]. All report significant improvement in the ability of BVI persons to navigate indoor spaces independently. However, the biggest challenge with BLE-based localization in GPS-limited areas is the scale of infrastructure deployment needed, making it practically infeasible for outdoor environments.

Localization without any infrastructure has relied on image processing/computer vision, or utilizing Wi-Fi access points. Limitations in Wi-Fi based localization include the need for a high density of access points to achieve acceptable accuracies for navigation, especially infeasible in outdoor environments. There has been much work done on image-based localization in the literature though there are very few works reported aimed for BVI persons. Thoma et al. proposed methods for image retrieval-based navigation in [10]. A CNN has been implemented in PoseNet [11] to relocalize a camera's 6-DOF pose relative to a scene. Clark et al. proposed a deep spatio-temporal model "VidLoc" in [12] to localize video-clip. All these works did not focus on requirements necessary for guiding BVI individuals in their outdoor wayfinding.

In order to assist BVI persons in their wayfinding, researchers used image-based localization by following different approaches. Tatsuya et al. [13] proposed such an approach combining structure from motion (SfM) and radio wave information from BLE beacons. Tatsuya et al. proposed another image-based localization in [14] by combining deep CNN with bluetooth radio-wave signal readings.

In all prior works, there has been little focus on what areas of outdoor pedestrian navigation have been troubling for BVI individuals and how image-based localization can provide much needed assistance to them. PathLookup, being a tagless system, is expected to be scalable by not needing infrastructure deployments. PathLookup also differs from prior work in being a unified approach, combining path prediction by a CNN network and other necessary location information from a contextual database to assist BVI persons in their outdoor wayfinding.

III. Motivation and Problem Formulation

A. A user experience study of relying only on GPS

A simple user study was conducted to identify the challenges of real-time outdoor wayfinding and to gather feedback from potential BVI individuals. Five participants with different levels of vision impairment were recruited for the study following appropriate institutional review board approvals. All participants were unfamiliar with the place where the test was conducted and were asked to find their way from a start point outside a building to a destination (338 ft away on the shortest route) (fig.1). The paths to the destination involved crossing a street. The location had good GPS coverage, and participants were given the option to use a mapping application of choice on their smartphone.

Table I shows the navigation time and navigation steps taken by the participants for the task. As can be seen, for a route that should take only about a minute to navigate, participants took much longer. Moreover, a significant variation can be seen in the number of steps taken by participants to reach the destination. Those who navigated, relied on Google Maps. However, Google Maps couldn't guide the users to navigate in the right direction at intersection while crossing a street and areas nearby the parking lot. These results clearly indicate that just using a GPS-based mapping application is inadequate

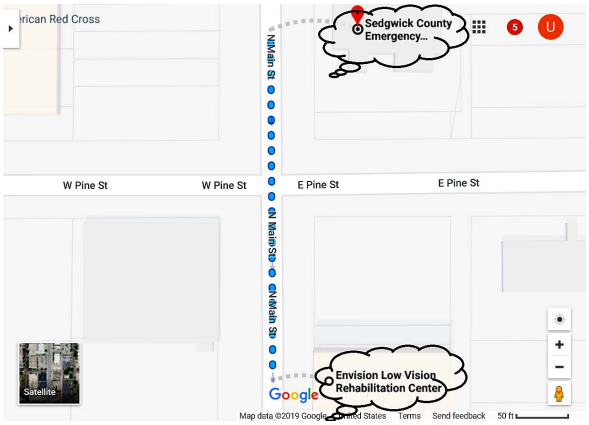

Fig. 1: A motivating Google Maps scenario

for the task of outdoor wayfinding, even when GPS coverage is good and pedestrian paths are adjacent to streets carrying vehicular traffic. As such, this study motivates the need of an auxiliary wayfinding system, to guide people with visual impairments in outdoor premises, complementing GPS-based mapping applications.

B. A quantitative study of GPS accuracy

Upon determining that Google Maps (utilizing underlying GPS signals) is not sufficient for outdoor navigation, a quantitative study of GPS accuracies for pedestrian navigation was conducted on another test site. Fig. 2 depicts the pedestrian navigation path in the test scenario, from Wallace Hall to Jabara Hall in Wichita State University, Wichita, KS, collected from Google Maps. This test site consists of many nearby buildings, trees, intersections and circles on the path, etc. As can be seen from fig. 2, Google Maps is unable to provide something close to the shortest path from source to destination; hence even a sighted person could be confused by Google Maps if this place is unknown to the person.

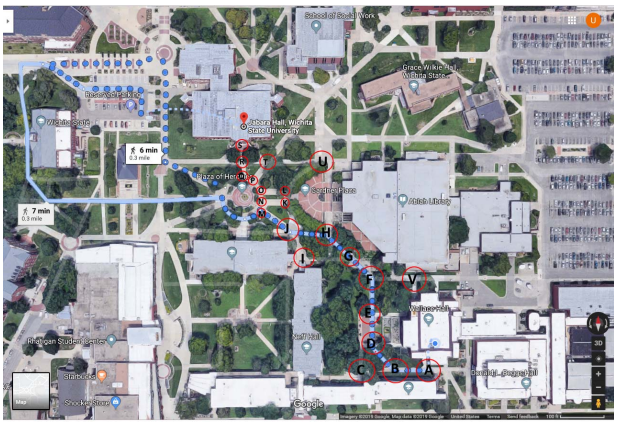

Fig. 2: Pedestrian navigation of a test scenario (PoI marked)

To investigate GPS coordinate accuracies received on this route (Wallace Hall to Jabara Hall), five different sets of data were collected for this study on a clear sky day. GPS

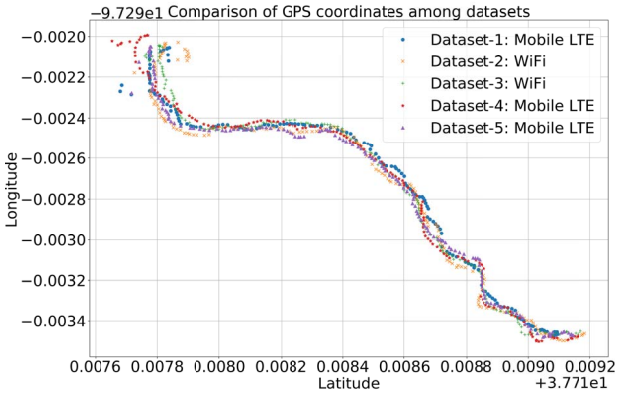

Fig. 3: Variation of GPS coordinates in a test site

coordinates in terms of longitude and latitude, and accuracy of the GPS were measured in this study. Due to the variability of the GPS coordinates (fig. 3) provided by the mapping application in the same route, a user could be confused to reach the destination properly and within the expected time. It was seen that for each navigation iteration, received corresponding GPS coordinates (in fig. 3) had different levels of accuracy. As such, any navigation application, developed by integrating Google Maps API on it and completely based on GPS coordinates in providing navigation guidance, would not be effective enough for a BVI person in outdoor wayfinding.

C. PathLookup Design Approach

Based on the user study and GPS accuracy study above, the motivation for an auxiliary outdoor wayfinding system such as PathLookup should be clear. Because computer vision techniques can perform localization without needing tags in the environment, it is the preferred approach in the design of the PathLookup system. CNN is more robust and reliable in unfavorable conditions such as motion blur, changing lighting conditions, etc. as opposed to local keypoint descriptor based approaches in outdoor wayfinding. Nevertheless, a deep-CNN based approach depends on extensive image training with high-dimensional features. Thus, high-dimensional feature extraction from training images and eventually predicting test images can be computationally expensive in terms of time and space if this type of approach is implemented real time on a smartphone. To mitigate storage and processing limitations of a phone, the solution approach used in PathLookup is that of training a deep-CNN based framework in the server, with end-users capturing an image from their smartphone and offloading it to the server for lookup while navigating, and finally receiving the results on the smartphone over the network.

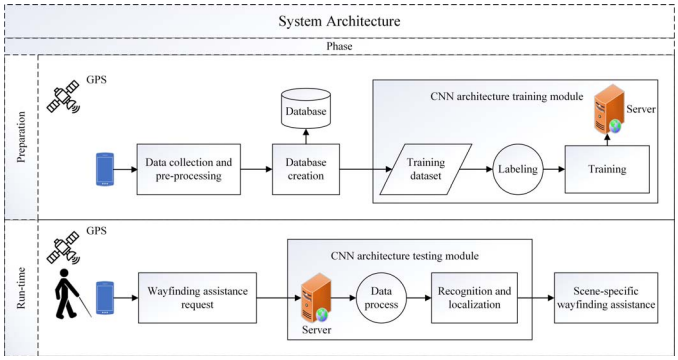

Fig. 4: System architecture of the PathLookup

IV. System Design

The PathLookup algorithm named "PLA-Sys" (Algorithm 1) is proposed in this section. The overall system architecture of PathLookup is illustrated in fig. 4 and demonstrated below:

A. Preparation phase

Data collection and pre-processing: A set of possible point of interests (PoI), γ was identified based on a candidate set, α of buildings / structures and a set of source-destination pair, β in a particular outdoor premise (fig. 2). Ground-truth GPS coordinates, σ are collected from Google maps and are used to label all PoI in the scene. All available directional paths ψ are created on each PoI in γ. Since β is already defined at the beginning of the "PLA-Sys" algorithm, clock orientation, ξ is identified for all available directional paths, ψ and thus used to label those elements.

Database creation: A contextual database, ∇ was created to provide critical information related to all identified directional paths. Among others, this contextual database contains available directional paths ψ on each POI, ground-truth GPS coordinate σ associated with each POI, clock orientation ξ, and path description λ. A directional path query to this database

TABLE I: User study

Participants	*Vision description*	*Time taken (minutes)*	*Steps needed (number)*	*Remarks / Feedback*
P-1	Cane user; Light perception (LP) on both eyes	4.2833	170	Used Google Maps before, feels that Google Maps isn't fine-grained enough for pedestrian navigation.
P-2	Cane user; LP on right eye, 20/500 on left eye	5.3833	282	Used Google Maps before for pedestrian navigation; thinks that it doesn't help much.
P-3	Guide dog user; No vision in one eye, LP on the other eye	8.25	404	Used Google Maps before but didn't help; Tried another app named "BlindSquare" as well for pedestrian navigation but didn't help either.
P-4	Visually impaired	N/A	N/A	Didn't feel comfortable using Google Maps; So didn't participate in testing by using Google Maps.
P-5	Visually impaired	N/A	N/A	Didn't feel comfortable using Google Maps; So didn't participate in testing by using Google Maps.

associated with path prediction, obtained by the computer vision part of the proposed algorithm (Algorithm 1), would be used in wayfinding assistance in the run-time phase of PathLookup.

Algorithm 1: PathLookup algorithm towards overall system development (PLA-Sys)

1 Initialize candidate set, α, of buildings / structures consists of potential source and destination,
$\alpha = \{b_1, b_2, b_3, ..., b_n\}$
2 Determine a set of source and destination pair,
$\beta = \{s, d\}$ s.t. $\beta \subsetneqq \alpha$
3 Create a set of possible point of interest (PoI), γ, within the area of the graph G(V, E), where V = α and E = connecting edges of α;
$\gamma = \{A, B, C, ..., X, Y, Z\}$
4 Label all elements of γ with ground-truth GPS coordinates, σ;
$\sigma = \{\sigma_1, \sigma_2, \sigma_3, ..., \sigma_n\}$ s.t. $f : \sigma \rightarrow \gamma$
5 Create a set of all available directional paths, ψ, on all particular PoIs s.t. $\psi \subseteq \gamma$;
$\psi = \{\{A_1, A_2, A_3, ..., A_n\}, \{B_1, B_2, B_3, ..., B_n\}, \{C_1, C_2, C_3, ..., C_n\}, ..., \{Z_1, Z_2, Z_3, ..., Z_n\}\}$
6 Label all elements of ψ with clock orientation, ξ, specific to the path of β;
$\xi = \{\xi_1, \xi_2, \xi_3, ..., \xi_n\}$ s.t. $f : \xi \rightarrow \psi$
7 Create a contextual database, ∇, containing path description $\lambda = \{\lambda_1, \lambda_2, \lambda_3, ..., \lambda_n\}$, s.t. $f : \lambda \rightarrow \psi$, along with other attributes;
$\nabla_i = \{\psi_i, \sigma_i, \xi_i, \lambda_i\}$,
where i denotes rows in ∇ and i = {1, 2, 3, ... , n}
8 Create an image dataset, X = $\{x_1, x_2, x_3, ..., x_n\}$ and label n samples (number of images) in X with ψ
9 Implement a CNN network for maximum probability prediction, Ω, of a path s.t. $\Omega \exists \psi$ and estimated by,
$\Omega_j = \arg\max_{j \in \psi} P(c_j \mid x_i)$,
where $P(c_j \mid x_i)$ is the probability of directional path class c_j given image x_i
10 Provide path advancement information based on path prediction by CNN and contextual database;
$\Theta_i = \{\Omega_i, \nabla_i\}$, where i = {1, 2, 3, ... , n}

Image training: A large image dataset, X corresponding to ψ was created and labeled appropriately in this work in order for the learning through the CNN model. If there is any overlapping among the elements of ψ along the path of β, the number of elements of ψ would be reduced by removing redundancy. The outcome of the CNN model would be the maximum probability prediction of a directional path.

B. Run-time phase

Wayfinding assistance request: During the run-time phase, while navigating in an outdoor premise and in a situation of uncertainty, a BVI user would capture an image from their smartphone camera on the route through the app. Through prior training, BVI users would be instructed beforehand on the appropriate orientation and height of the camera from which images should be captured. The captured image would then be offloaded to the server by the app for localization in a particular scene.

Recognition and localization: After a captured image/frame is received at the server, it would be tested on the CNN model and the result would be returned to the app on the smartphone over the network. The result would be in the format of Θ_i that is obtained by the proposed algorithm (algorithm 1) and could be integrated in any navigation app. This Θ_i could provide accurate scene-specific wayfinding assistance in route decision to BVI persons in a particular scene.

V. Image localization network

A. Dataset

Several high definition videos of ψ along the path of β = {Wallace Hall, Jabara Hall}, in Wichita State University campus, were captured by a Samsung Galaxy Note 8 smartphone as per algorithm 1. Then those videos were converted to images/frames at the resolution of 1080 x 1920 at 5 fps. Image dataset for localization network was created from three different sets of videos captured in different weather conditions and timings (table II). All videos were collected in real-time during the working day of the university. The camera of the smartphone was held at chest-height as a user may do when taking an image for lookup. A total of 20716 images are used in training dataset and a total of 5179 images are used in validation dataset for image localization network (table II).

TABLE II: Image dataset for CNN network

	Set 1	Set 2	Set 3
Weather	Rainy/cloudy	Sunny	Sunny
Timing	12 pm - 2 pm	10 am - 12 pm	2 pm - 4 pm
Total no. of images	8589	8485	8821
No. of training data	20716 (80% of total images)		
No. of validation data	5179 (20% of total images)		

B. CNN architecture model

A CNN architecture model (fig. 5) is proposed and implemented in this work for PathLookup in order to train the image dataset associated with ψ as per algorithm 1. The pre-processed images are labeled with appropriate ψ value and other necessary information. The CNN network is designed empirically and is implemented by using *Keras 2.0*, a deep learning Python library. All input images are resized to the resolution of 128 x 128 and only one channel (gray-scale) was used in the training to reduce the computational complexity. The proposed CNN network is composed of a stack of convolutional modules that perform feature extraction. A final fully connected layer containing number of ψ units is created for the image localization network. The proposed CNN model is trained to output a probability over ψ directional path classes for each given image.

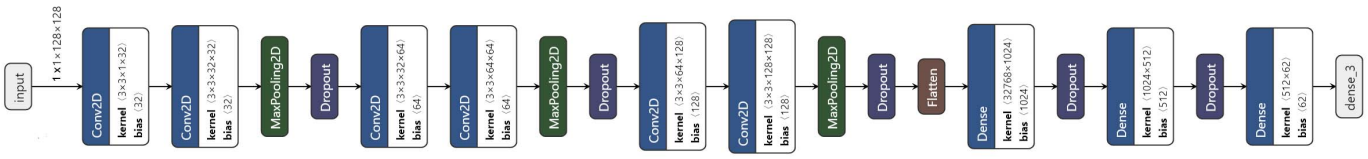

Fig. 5: Architecture of the proposed CNN network of PathLookup

VI. Experimental results

A. Methodology

The methodology used to evaluate our CNN model includes "model accuracy" and "classification performance". Through these metrics it is understood that how the proposed model in this work performs in identifying a path in pedestrian navigation. These information would be very useful in pedestrian navigation for BVI persons. A commodity laptop computer (Intel Core i5-4200U, CPU @1.60 GHz, 8.0 GB RAM, and Intel HD Graphics) was used to train the CNN network in the experiment.

B. Results

(a) Loss curve

(b) Accuracy curve

Fig. 6: Loss and accuracy curve comparison

1) Model accuracy: Fig. 6 illustrates the loss and accuracy curves of the model comparing between training and validation dataset. As can be seen from the fig. 6, the plot of training loss and validation loss decreases to a point of stability after epoch 15 and also validation loss has a small gap with the training loss. These indicate our model has good fit learning curves. Moreover, at the end of epoch 40, the model achieved 97% of training accuracy and 99.52% of validation accuracy.

2) Classification performance: Fig. 7 depicts the classification performance of our model for predicting a directional path in the test site. In this plot, "Precision" defines the quantity of how much our model predicts correctly out of all the directional path classes in ψ. As can be seen from the plot (fig. 7), the value of "Precision" is 1 in 50 directional path classes out of 62, while it varies in the range from 0.94 to 0.99 in rest of the classes. Thus, a higher value of "Precision" was achieved for all directional path classes in ψ. Moreover, how much our model predicts correctly, out of all the positive classes, is defined by "Recall". Fig. 7 shows the value of "Recall" is 1 in 48 classes out of 62. In rest of the classes, the "Recall" value varies from 0.90 to 0.99. Overall, a higher value of "Recall" was also achieved for all directional path classes in ψ. The values of "F1-score" (harmonic mean of "Precision" and "Recall") are also illustrated in fig. 7 as compared to "Precision" and "Recall" values.

C. Discussion of results

The proposed CNN model of PathLookup achieved 97% accuracy on the training dataset. Since the training dataset was prepared based on real-time images collected during the business days of the University, pedestrians and other objects located along the testing path created noise and did impact on the training accuracy. However, still, the CNN model of the PathLookup framework could predict an image accurately enough (99.5%) along the path of source-destination on the validation dataset. The classification error incurred by the CNN network happened due to the lack of enough distinction of features in closely spatially related directional paths. Since the PathLookup framework could predict an image accurately enough along the path of source-destination and also provides critical information of a particular path, the proposed framework will be suitable to be integrated into any real-time navigation application for blind and visually impaired users.

In addition to being accurate, PathLookup is found to make an image prediction within the order of a few milliseconds utilizing only a commodity laptop computer as the back end server. By utilizing a server's resources through an offloading paradigm, the storage, battery, and processing limitations of smartphones are circumvented.

VII. Conclusions and Future work

Through a user study and quantitative study of GPS accuracies, this paper demonstrated the need for a mechanism to assist BVI persons in outdoor wayfinding complementing existing GPS-based systems. Subsequently, a deep learning-based computer vision framework named PathLookup was

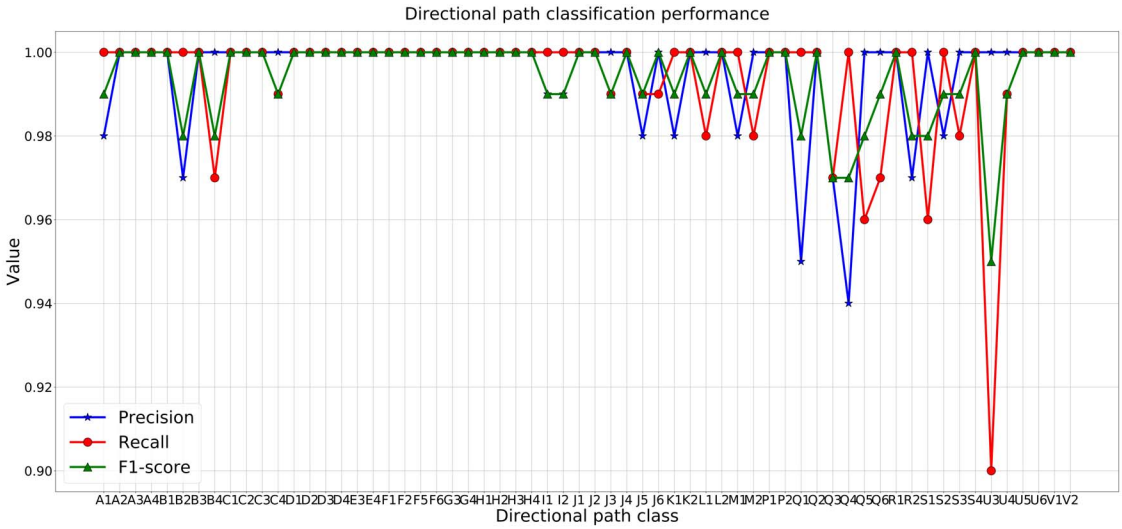

Fig. 7: Classification performance of all directional path classes

presented for image localization that can help with outdoor wayfinding for BVI individuals. The experimental results show that the CNN network of PathLookup performs with 99.5% accuracy on the validation dataset in predicting the path toward a destination. The outcome of the CNN network in conjunction with other location information from a contextual database as constructed PathLookup can thus prove to be of great assistance to BVI persons for outdoor wayfinding. This work has made only the technical contributions towards an image localization tool for outdoor wayfinding of BVI persons; there remains scope for much additional work before it will be ready to use as a commodity application for BVI individuals.

ACKNOWLEDGEMENT

This work was funded in part by the U.S. National Science Foundation (NSF) CNS # 1951864.

REFERENCES

[1] P. Zandbergen and S. Barbeau, "Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones," *Journal of Navigation*, vol. 64, pp. 381 – 399, 07 2011.

[2] J.-E. Kim, M. Bessho, S. Kobayashi, N. Koshizuka, and K. Sakamura, "Navigating visually impaired travelers in a large train station using smartphone and bluetooth low energy," in *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, ser. SAC '16, 2016, pp. 604–611.

[3] D. Ahmetovic, C. Gleason, C. Ruan, K. Kitani, H. Takagi, and C. Asakawa, "Navcog: A navigational cognitive assistant for the blind," in *International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2016.

[4] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, "Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment," in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. ASSETS '17, 2017, pp. 270–279.

[5] S. A. Cheraghi, V. Namboodiri, and L. Walker, "GuideBeacon: beacon-based indoor wayfinding for the blind, visually impaired, and disoriented," in *IEEE International Conference on Pervasive Computing and Communications*, March 2017.

[6] R. Manduchi and J. Coughlan, "(computer) vision without sight," *Commun. ACM*, vol. 55, no. 1, pp. 96–104, Jan. 2012. [Online]. Available: http://doi.acm.org/10.1145/2063176.2063200

[7] R. Jafri, S. A. Ali, H. R. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: A survey," *Vis. Comput.*, vol. 30, no. 11, pp. 1197–1222, Nov. 2014. [Online]. Available: http://dx.doi.org/10.1007/s00371-013-0886-1

[8] V. Nair, M. Budhai, G. Olmschenk, W. H. Seiple, and Z. Zhu, "ASSIST: Personalized indoor navigation via multimodal sensors and high-level semantic information," in *The European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[9] A. G. et. al., "PERCEPT navigation for visually impaired in large transportation hubs," *J. Technol. Persons Disabilities*, pp. 336–353, March 2018.

[10] J. Thoma, D. P. Paudel, A. Chhatkuli, T. Probst, and L. V. Gool, "Mapping, localization and path planning for image-based navigation using visual features and map," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7375–7383.

[11] A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 2938–2946. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2015.336

[12] R. Clark, S. Wang, A. Markham, N. Trigoni, and H. Wen, "Vidloc: A deep spatio-temporal model for 6-dof video-clip relocalization," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2652–2660.

[13] T. Ishihara, J. Vongkulbhisal, K. M. Kitani, and C. Asakawa, "Beacon-guided structure from motion for smartphone-based navigation," in *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2017, pp. 769–777.

[14] T. Ishihara, K. M. Kitani, C. Asakawa, and M. Hirose, "Deep radio-visual localization," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2018, pp. 596–605.