Does your robot know when to cross the road?

Muneeb Shafique
Department of Computer Science
Habib University
Karachi, Pakistan
ms06373@st.habib.edu.pk

Abdul Majid
Department of Computer Science
Habib University
Karachi, Pakistan
at06616@st.habib.edu.pk

Sajeel Alam
Department of Computer Science
Habib University
Karachi, Pakistan
sa06840@st.habib.edu.pk

Abdul Samad
Department of Computer Science
Habib University
Karachi, Pakistan
abdul.samad@sse.habib.edu.pk

***Abstract*—This paper presents a novel approach to enhance a robot's proficiency in correctly identifying Pedestrian Traffic Light (PTL) signs in adverse conditions. The pivotal aspect of this endeavor lies in the accurate recognition and interpretation of the pedestrian traffic light signals. By discerning signals such as the red pedestrian icon, countdown timer, or absence of display, the robot can make informed decisions to abstain from crossing. Conversely, upon detecting the green pedestrian icon, the robot can safely traverse the road. The primary objective of this study is to improve a model capable of processing images taken from the perspective of pedestrians, providing precise outputs corresponding to the exhibited PTL sign and improve its accuracy in unfavourable conditions. Considering the global variability in PTL designs, our focus is tailored to the specific street conditions prevalent in Japan. By enhancing the data the model is currently trained on, this research strives to enhance the robotic system's adeptness in recognizing and responding to PTL indications within Japanese crosswalks.**

***Index Terms*—Deep-learning, Neural networks, Data set, Traffic Light, Zebra-crossings**

I. Literature Review

Over the past few decades, robots have become an integral part of our society from being used for automation in factories to exploring other planets in the form of space rovers. Robotic technology is used to provide lost limbs to amputees and as well for military defense systems. There are plenty of use cases for robots however, the one we are studying in this project are socially aware or autonomous delivery robots and training a model that can guide them to safely cross roads.

After extensive research, we realized that there are multiple studies on navigation systems and hazard detection for the visually impaired in order to help them become more independent in commuting within cities. In this domain, there exist both sensor based and deep learning based approaches. Some works such as the MOVIDIS Project include using a radio frequency based communication system to help visually impaired people use the public transport system [1]. Others include using ultrasonic sensors and convolutional neural network to construct a wearable device that detects obstacles such as potholes and then text to speech is used to warn the user [2]. Context-Aware Assistive Systems (CAAS) are used widely in autonomous driving however, combined with machine learning it has been used to develop smart glasses for hazard detection [3].

Although these approaches may sound revolutionary, it is important to note that hardware is usually expensive and also has slow response times that fail in real life situations. Hence, it is essential to explore deep learning models that can be used navigation systems. Previous studies suggest that such models have been trained for traffic light detection in autonomous driving however, these lights are always circular and have no variations globe thus they do not pose much of a challenge [4]. If we narrow down our problem from navigation systems to solely pedestrian traffic light (PTL) detection, a few studies to can be found where models are trained to identify red and green PTLs in real time. One such study has developed a detection algorithm that uses candidate extraction, candidate recognition and temporal-spatial analysis to differentiate between red and green PTLs. An SVM model was trained on data from China, Italy and Germany where images from the pedestrian's point of view were provided. The model performed really well except in rainy conditions where it had a poor recall [5].

All the research discussed up till now has been for aiding visually impaired people but it is essential to understand that research for navigation systems for the visually impaired overlap robot navigation. A particular study divides this problem into three parts where it analyses PTL detection and crosswalk detection in environment mapping, route selection in journey planning and avoiding obstacles in real time navigation [6]. These three challenges are common to both visually impaired people and robots hence, whatever work has been done for guiding visually impaired people to cross roads can also be applied to robots.

Prior to solving a problem, it is essential to establish the

979-8-3503-1770-1/23/$31.00 ©2023 IEEE

need for it. So we ask ourselves why are autonomous robots important? Previous studies build a strong argument where one in particular talks about Last-Mile Delivery concept using robots. In this concept, a delivery truck delivers packages to a micro-depot from where assigned robots deliver packages to houses. This reduces traffic, pollution and increases efficiency for delivery companies [7]. In order for a robot to deliver packages, it is vital for it have excellent navigation skills so that it does not pose a threat to other pedestrians and vehicles. A study mentions using robot operation system (ROS) to carry out accurate robot navigation that is aimed to solve the SLAM (Simultaneous Localization and mapping) problem [8]. In this model, the robot localizes itself within a map and traverses a path on its own. The study showed promising results but it would require a lot of tuning and trial and error for various robots in different situations. Moreover, this concept is for simple robot navigation and does not take into consideration PTL detection and real life challenges that the robot might encounter while crossing roads.

Going through all the previous studies and analyzing the work done assisted us in specifying our problem statement to robots knowing when to cross the road by correctly identifying PTLs for the streets of Japan. A particular study which we decided to improve upon is one in which a model named LYTNetV2 has been trained to detect and identify PTLs for the visually impaired. The model was trained on data from Shanghai where images were provided form the pedestrian's point of view and the output can be from five classes (red, green, countdown green, countdown blank, none) [9]. The accuracy of 94% is impressive which we plan on improving for data from Japan by altering the model and training it on categories that it performs poorly in.

II. Problem Statement

The problem at hand pertains to the limited accuracy of current pedestrian traffic light (PTL) recognition systems, particularly when faced with non-ideal conditions characterized by inclement weather (such as rain or snow), excessive sunlight, blurriness, or the presence of distracting elements in the images. These challenging conditions accentuate the severity of the problem at hand. Problems of this nature have already been detected in parallel areas such as the use of specialized models to improve sign detection accuracy [10]. Inaccurate PTL detections in such circumstances can lead to potentially hazardous situations for pedestrians and traffic, highlighting the compelling need for a solution that not only elevates accuracy under optimal conditions but, critically, excels in addressing these adverse environmental challenges.

III. Research Methodology

The primary objective of this study was to enhance the accuracy of our model in detecting PTL (Pedestrian Traffic Light) signs. Achieving this goal demanded a systematic approach that focused on both the quantity and quality of the dataset i.e. largest open access PTL data set provided by Samuel Yu and Heon Lee. This data set comprises of 5219 images of junctions captured in Shanghai, China. To effectively address these aspects, our research methodology was divided into three key steps:

1) **Data Analysis and Categorization**: This step revolved around improving the quality of the dataset. We undertook a comprehensive analysis aimed to understand the strengths and limitations of the existing dataset. We defined categories based on the features and characteristics of the data, allowing us to classify the dataset into these distinct categories. In addition to categorizing the existing data, we identified new categories that addressed the limitations in the original dataset. We considered the types of images that were missing from the dataset but could potentially be encountered in real-world situations. This process was crucial for enhancing the model's performance in scenarios it had not been exposed to during training.
2) **Data Augmentation**: To bolster the quantity of data, we proceeded to acquire additional images for both the categories present in the original dataset and the newly identified categories. This step involved an extensive data augmentation process, which included the collection of images that were reflective of the identified categories.
3) **Model Evaluation:**:The final step of our research methodology involved rigorous testing to evaluate whether the model's accuracy had genuinely improved following the enhancements to both data quantity and quality.

IV. Research Questions

- How can the accuracy of Pedestrian Traffic Light (PTL) recognition systems be improved in adverse environmental conditions, such as rain, snow, and strong sunlight, to enhance their reliability for real-world robot applications?
- Can the performance of PTL recognition models be enhanced by collecting and incorporating a comprehensive dataset of Japanese crosswalk scenarios, accounting for variations in PTL design and urban environments?
- How does the accuracy and reliability of PTL recognition systems impact the overall safety of autonomous robotic systems when crossing streets?

V. Data Analysis and Categorization

A short description of each category and their count in the original dataset is given below.

A. Green PTL and blocked by Car

This category highlights those instances where Cars drive along the crosswalk despite the Pedestrian Light being Green. It indicates those examples where looking at only the PTL for decision making is not adequate. An example image is shown below:

Fig. 1. The data set includes 237 images of this type

B. 2 or more PTLs

This category highlights those instances where there are more than one PTL in the captured image. In this case, the model itself has to decide which of the one of the PTL is relevant in crossing the junction. An example image is shown below.

Fig. 2. The data set includes 135 images of this type

C. No PTL with crosswalk

This category highlights those instances where the image has no PTL, despite there being a cross walk. This could be due to the angle of the image or there being no PTL installed by the city. An example image is shown below.

Fig. 3. The data set includes 382 images of this type

D. Pedestrian Lines Blurred

This category highlights those instances where the crosswalk lines were light or completely erased making determination of start point and end point tougher. An example of such an image is shown below.

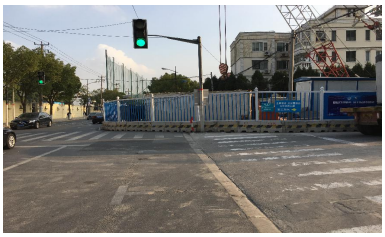

Fig. 4. The data set includes 106 images of this type

E. PTL Of Irrelevant Junction shown

This category highlights those instances where the image displays only one PTL but of a junction that the person was not intending to cross.An example of such an image is shown below.

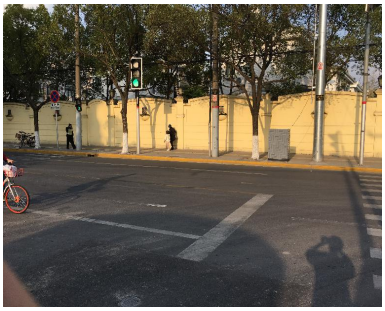

Fig. 5. The data set includes 7 images of this type

F. Low Light

This category highlights those instances where background light is dim (such as evening time images or images taken under shade of a bridge). An example of such an image is shown below.

Fig. 6. The data set includes 73 images of this type

G. Bad Lighting On PTL

This category highlights those instances where the color of the PTL is not easily discernible due to glaring sunlight on the PTL or due to dimness in the PTLs own light. An example of such an image is shown below.

Fig. 7. The data set includes 94 images of this type

H. PTL not working

This category highlights those instances where the there is a PTL light but the light is blank and does not show any of the expected signs. This could be due to frame rates of the video used to extract the images or due to technical fault in the PTL. An example of such an image is shown below.

Fig. 8. The data set includes 62 images of this type

I. Sun Pointed towards camera

This category highlights those instances where the sun rays beam right at the lens of the camera which has distorting effects on the image.

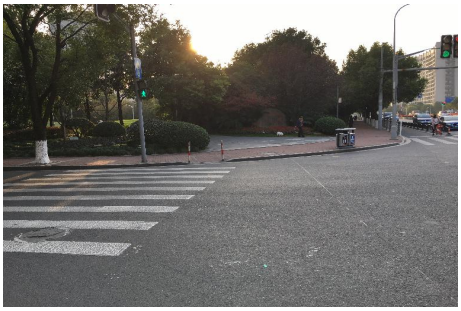

Fig. 9. The data set includes 14 images of this type

J. PTL too far

This category highlights those instances where the PTLs were too far away from the camera. This generally occurs in larger intersections and decreases the size and hence visibility of the PTL. An example of an image in this category is shown below.

Fig. 10. The data set includes 50 images of this type

K. Blurred Images

This category highlights those instances where the image used for determination of the PTL sign were blurred. An example of an image in this category is shown below.

Fig. 11. The data set includes 32 images of this type

L. PTL obscured

This category highlights those instances where the PTL is blocked by cars or other obstacles. This category contains images for both partial obstruction and complete obstruction. An example of complete obstruction in an image is shown below.

Fig. 12. The data set includes 25 images of this type

M. Rain

This category contains images taken in or just after it has rained. One such example is shown below.

Fig. 13. The data set includes 144 images of this type

N. Distracting signs

This category contains images that along with the PTL have a lighted signboard featuring the same colors. This can confuse the model as it could read the color on the signboard and make a prediction.

Fig. 14. The data set includes 76 images of this type

O. Red and Green on the same PTL

This category features images where a faulty PTL displays both red and green man signs simultaneously. A example of such an image is shown below.

Fig. 15. The data set includes only 1 image of this type

P. Snow

This category features images that were taken when it was snowing. Falling snow obliterates the PTL on many occasions affecting visibility of the PTL

Fig. 16. The data set includes 48 images of this type

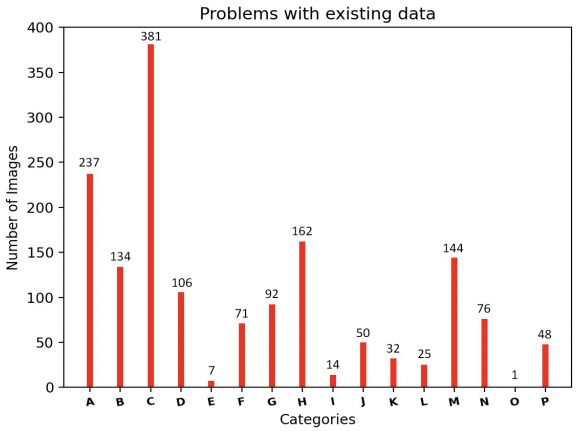

Fig. 17. Graph summarizing number of images within each identified category

VI. Data Collection & Augmentation

A. Categories:

From the above identified 16 categories (A-P), we selected the following 10 categories and began data collection for them.

- (G) Bad Lighting on PTL
- (K) Blurred Images
- (N) Distracting Signs
- (C) No PTL with Crosswalk
- (H) PTL not working
- (L) PTL obscured
- (J) PTL too far
- (P) Snow
- (B) 2 or More PTLs
- (F) Low Light

6 categories were dropped during this phase as it was difficult to obtain data for those categories, or the model performed well on the challenging categories. For example, category A (Green PTL blocked by a car.) During data collection, we observed that Japanese citizens are law-abiding. It is extremely

rare to find Car drivers breaking traffic signals. Contrast this with China, multiple instances were found of cars crossing regardless of the lights. Similarly, we noticed that category D (blurred pedestrian lines) did not impact the model's performance, as indicated by Fig 18. Furthermore, category N (Rain) did not prevent the model from accurately predictin the color of the Pedestrian traffic lights. For Category E (PT of irrelevant junction shown) there is no way for the model t determine whether the PTL in the image belongs to its junctio or some other junction. In the case of category I (Sun pointe towards the camera), due to the low exposure the PTL become easier to identify thus making this an ideal condition. Categor O (Red and green on the same PTL), was encountered whe the image was captured at the exact frame where the signs ar changing. In the case of a robot in the real world, an imag taken a few frames later will not have this issue.

B. Data Collection:

We used the following three resources to collect data fo the categories identified in the above section.

- Youtube
- Videos (Mr. Zulkafil Abbas)
- Data Augmentation

Numerous YouTubers in Japan take their subscribers on live walks for several hours. One such Youtuber that proved to be super useful for us was Tokyo Explorer. During these long walks, they come across junctions of all sorts, which provided us with an opportunity to take screenshots of Pedestrian traffic lights and subsequently categorize them. Although this was a painstakingly slow process, we obtained high-resolution images with this source. Also, with the help of our supervisor Mr. Zulkafil Abbas who is located in Japan, we were able to fill the gaps in our data for categories not addressed by the Youtube videos. This was a great resource as it allowed us to dictate how the videos were taken of different Pedestrian traffic lights. A youtube channel (PTL Dataset) was set up for the exchange of such videos, and a python script was used to extract image frames from them. Each extract was then either placed into one of the categories or discarded to ensure that no more than two images were taken of the same Traffic light. Lastly, for categories that are seasonal in nature, such as Snow, we used Photoshop to apply filters to already collected images. Similarly, editing tools in Photoshop, such as increasing exposure, were used to simulate the effect of category G (Bad lighting on PTL). Using these sources, we were able to collect a total of 1072 useful images for the identified categories.

VII. Testing Accuracy of Existing Model w.r.t Identified Categories

Accuracy is perhaps the best-known Machine Learning model validation method used in classification problems. To improve the accuracy of the exisiting model, it is important to check the model's performance on the identified categories. To do this, the model is used to predict answers on images from the individual categories which is then compared to ground truth values.

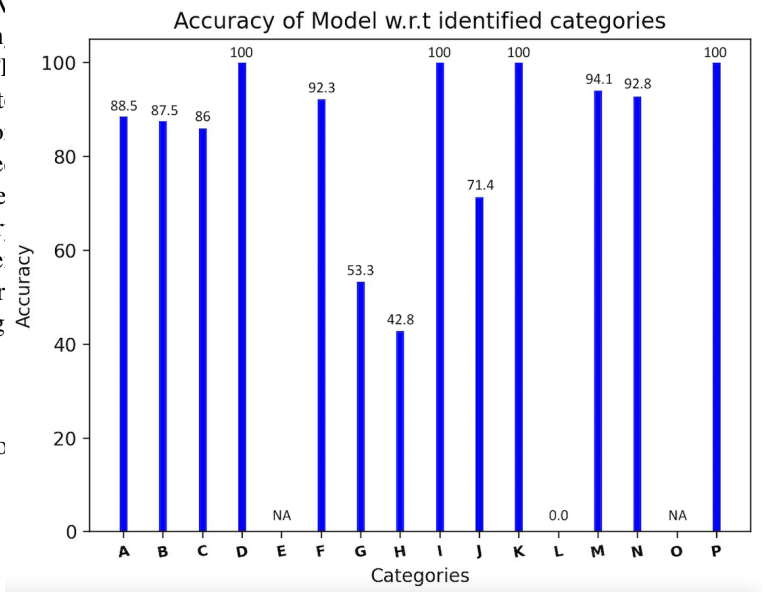

Fig. 18. Performance of Model w.r.t to Identified Categories

The purpose of carrying out this accuracy exercise is to improve the performance of the existing model. The graph highlights the models' comprised performance when there is excessive sunlight on the PTL (53.3%) or when the PTL is too far away (71.4%). To improve the performance, we will feed the existing model with increased training examples specific to these challenges so as to familiarize the model with these situations.

VIII. Model and Experiments

A. Model Details:

For our work to be practical and useful in real life scenarios, we needed to use a model that could be easily applied to handheld devices and still provide great accuracy. The model we have used is LYTNetV2. LYTNetV2 is derived from MobileNetV2 which as its name suggests is purpose built as a light Convolutional Network for mobile phones. Like MobileNetV2, LYTNetV2 also uses depthwise separable convolutions so each input channel is seperately treated and different filters are applied on it. In order to keep the parameters low and computation quick, LYTNetV2 uses several bottle neck layers, these layers help the network compress the feature representations to best fit available space. In order to recognize the difficult category images we have just mentioned, there is a need for higher resolution images that capture greater detail hence LYTNetV2 is perfect since it has expanded MobileNetV2's input range to have 768 x 567 sized inputs. The Network uses lesser number of bottlenecks as compared to MobileNetV2 in order to capture greater detail. After the depth wise convolutions the model also uses fully connected layers

in order to produce the results. LYTNetV2 also introduces use of squeeze excite layers which allows the model to focus on the most important features of the images while weighing the less important ones. Historically LYTNetV2 has obtained best possible accuracy in PTL based applications and hence forms the model of choice for our application. As indicated by [4], the first LYTNet provided 92% Accuracy in previous attempts to solve the problem. A detailed view of the network architecture is shown below:

TABLE I
LYTNETV2 NETWORK DETAIL

Input	Operator	k	c	SE
768 x 3	conv2d	3	16	No
384 x 16	maxpool	2	-	No
384 x 16	bottleneck	3	16	No
192 x 16	bottleneck	3	24	No
96 x 24	bottleneck	3	24	No
96 x 24	bottleneck	5	40	Yes
48 x 40	bottleneck	5	40	Yes
48 x 40	bottleneck	3	80	No
24 x 80	bottleneck	3	80	No
24 x 80	bottleneck	3	112	Yes
24 x 112	bottleneck	5	160	Yes
12 x 160	bottleneck	5	160	Yes
12 x 160	bottleneck	3	320	No
12 x 320	conv2d	1	960	No
12 x 960	avgpool	-	-	No
1^2 x 960	conv2d	1	1280	No
1280	FC	-	5,4	No

[a]SE stands for Squeeze Exite Layer

B. Re-Training the Model:

To retrain our model, we first divided the 1072 obtained images into training, testing, and validation datasets using the ratio 8:1:1. This left us with 831 additional images in the training dataset, 120 in the validation dataset, and 121 in the testing dataset. All these images were then individually labeled for retraining and testing the model. Since these images were taken from various sources, they had varying resolutions and had to be preprocessed for the model to accept them. To deal with random cropping, the training images were set to a resolution of 876 x 657. On the other hand, validation and testing images did not have to be randomly cropped. Hence they were set to a resolution of 768 x 576. The additional 831 training images were then added to the original training dataset, thus giving us a total training dataset consisting of 4287 images, whereas the additional 120 validation images were added to the original 864 validation dataset, thus giving us a total validation dataset consisting of 984 images. These images were then fed into the model, which was retrained keeping the hyperparameters the same as before. After the model was retrained, we tested it for each of the selected categories and got our results.

IX. RESULTS AND DISCUSSION:

A. Results

As mentioned earlier, our goal was to focus on collecting data for the 10 selected categories, retrain the model on the updated data and then test the model again to check whether the testing accuracies have improved. Hence we present our results:

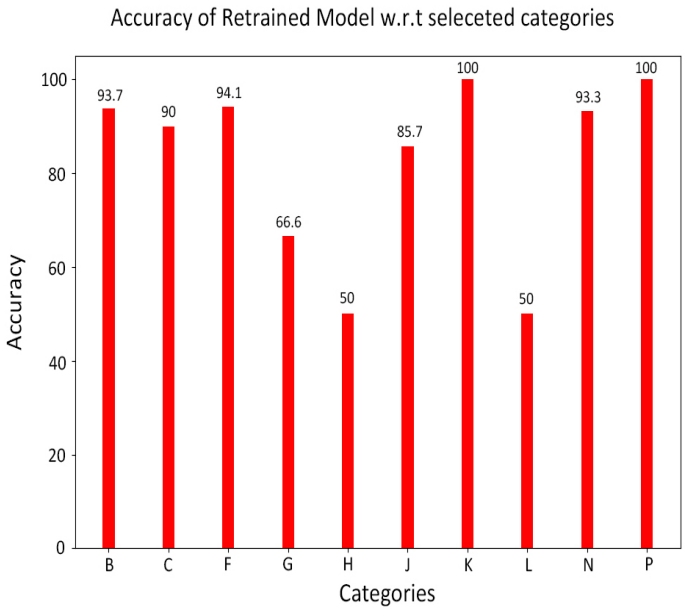

Fig. 19. Performance of Retrained Model w.r.t to Selected Categories

Comparing Fig 18 and Fig 19, we can see that the accuracies have generally improved. This is represented more clearly in the table below:

Category Name	Old Accuracy	New Accuracy
2 or more PTLs (B)	87.5	93.7
No PTL with crosswalk (C)	86	90
Low Light (F)	92.3	94.1
Bad Lighting On PTL (G)	53.3	66.6
PTL not working (H)	42.8	50
PTL too far (J)	71.4	85.7
Blurred Images (K)	100	100
PTL obscured (L)	0	50
Distracting Signs (N)	92.8	93.3
Snow (P)	100	100

Fig. 20. Comparing Performance of Model and Retrained Model

For category G (Bad Lighting On PTL), we expected the accuracy to improve more than it actually did since we added 98 additional challenging images in its training dataset. However, categories B (2 or more PTLs) and J (PTL too far) showed promising results as their accuracies increased from 87.5 to 93.7 and from 71.4 to 85.7, respectively. In the future, we can further work on the categories that did not show much improvement by gathering more challenging data specific to that category and retraining the model.

B. Future Directions:

Attempting to establish aids for individuals who are visually impaired has urged many cities to seek solutions for improving their quality of life. Namely, cities have installed sound-emitting devices into traffic lights and sidewalks that assist their navigation. Our model has the potential to help in this regard if it can be integrated into a mobile application since it achieves a high accuracy regardless of daylight variability due to time and weather. To make this algorithm effective and truly useful for everyday use, a better running time is required (Less time-consuming functions). In addition, a solution for twilight time is required. This project can be expanded to other road cross themes, such as Zebra Crossing Detection. Furthermore, this existing model can be used in red light cameras to enforce traffic laws.. These cameras are located at busy intersections to detect cars/motorists breaking red lights. To accurately process this violation, it is essential that the camera correctly detects the color of the traffic lights and then captures an image of the vehicle.

REFERENCES

[1] Y. Sáez, J. G. Parera, F. Canto, A. García González, and H. Montes, "Assisting visually impaired people in the public transport system through rf-communication and embedded systems," *Sensors*, vol. 19, pp. 10–12, 03 2019.

[2] M. M. Islam, M. S. Sadi, and T. Bräunl, "Automated walking guide to enhance the mobility of visually impaired people," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 3, pp. 485–496, 2020.

[3] O. Younis, W. Al-Nuaimy, F. Rowe, and M. H. Alomari, "A smart context-aware hazard attention system to help people with peripheral vision loss," *Sensors*, vol. 19, no. 7, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/7/1630

[4] S. Yu, H. Lee, and J. Kim, "Lytnet: A convolutional neural network for real-time pedestrian traffic lights and zebra crossing recognition for the visually impaired," in *Computer Analysis of Images and Patterns*, M. Vento and G. Percannella, Eds. Cham: Springer International Publishing, 2019, pp. 259–270.

[5] R. Cheng, K. Wang, K. Yang, N. Long, J. Bai, and D. Liu, "Real-time pedestrian crossing lights detection algorithm for the visually impaired," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 20 651–20 671, Aug 2018. [Online]. Available: https://doi.org/10.1007/s11042-017-5472-5

[6] F. E.-z. El-taher, A. Taha, J. Courtney, and S. Mckeever, "A systematic review of urban navigation systems for visually impaired people," *Sensors*, vol. 21, no. 9, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/9/3103

[7] M. Poeting, S. Schaudt, and U. Clausen, "A comprehensive case study in last-mile delivery concepts for parcel robots," in *2019 Winter Simulation Conference (WSC)*, 2019, pp. 1779–1788.

[8] H. Anas and O. W. Hong, "An implementation of ROS autonomous navigation on parallax eddie platform," *CoRR*, vol. abs/2108.12571, 2021. [Online]. Available: https://arxiv.org/abs/2108.12571

[9] S. Yu, H. Lee, and J. Kim, "Street crossing aid using light-weight cnns for the visually impaired," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 2019, pp. 2593–2601.

[10] X. Gao, L. Chen, K. Wang, X. Xiong, H. Wang, and Y. Li, "Improved traffic sign detection algorithm based on faster r-cnn," *Applied Sciences*, vol. 12, no. 18, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/18/8948