International conference on Advanced Techniques in Communication Networking and Automation

RESEARCH ARTICLE OPEN ACCESS

A Prototype of Object Recognition and GPS System for Partially Sighted Humans using CNN

T. Gurupriya*, S.Vignesh**, A.R. Arjun***, R. Manmadhan****,
S. Adrian Supreme Kumar *****
Department of Electronics and Communication
Periyar Maniammai Institute Of Science and Technology, Vallam -613403
Email Id: gurupriya@pmu.edu*, vigneshselvarasu120@gmail.com**, arjunar2468@gmail.com***, manmadhanr1442000@gmail.com****, adriansupremekumar10@gmail.com *****

ABSTRACT
Technology has enabled innovative devices to assist individuals in their daily lives, including visually impaired persons (VIPs). To address the mobility and safety needs of VIPs, a smart and intelligent system has been designed that provides real-time navigation using an automated voice and a web-based application for location sharing with family members. The system allows VIPs to sense and visualize their surroundings and provides security through location tracking. Unlike existing systems that require separate microcontrollers for different operations, this system uses Mobile Net architecture with low computational complexity to run on low-power end devices.

I. INTRODUCTION

Object detection is a crucial part of computer vision, allowing for the location and classification of objects in an image. Real-time object detection is essential for applications such as autonomous vehicles and augmented reality devices. However, current state-of-the-art algorithms require expensive sensors or significant computational power. YOLO-LITE is a lightweight algorithm that aims to provide real-time object detection on standard computers.

This paper presents a smart system that performs real-time object localization and recognition for visually impaired individuals, providing audio feedback for recognized objects and tracking the user's location for family members. The system uses mobileNet architecture for low computational complexity on wearable devices.

II. EXISTING SYSTEM

The existing system uses separate microcontrollers for different operations such as object detection and emergency call using GSM. Another system called "Let Blind People See" uses audio feedback to alert users of obstacles in real-time. The device uses a mobile camera and the YOLO algorithm for object detection and creates 3D locations and audio for objects through a 3D game-based application. The device works in both indoor and outdoor environments but has difficulty with heavy computation and detecting objects at longer ranges. Testing on eight individuals, three of whom are blind, showed the system is useful in unknown environments. The device provides continuous instructions to keep the user on track and has 97.33% accuracy for obstacle detection at 100cm.

III. PROPOSED SYSTEM.

The proposed system includes a DSP board, GSM and GPS modules, headphones, and a camera. The DSP board captures the live video feed and passes it to the object detection and recognition module, which predicts the objects in the video frame. The text to speech converter pronounces the object names through headphones, and the labelled snapshot with the user's location is saved in a server database. A user-driven feature allows VIPs to share their location with family members through a web interface, which can be enabled or disabled with a single button press.

A) DSP BOARD

Fig 1-DSP BOARD

The Raspberry Pi 3 Model is a small on-board computer developed by the Raspberry Pi Foundation. It has a system on a chip with an integrated ARM RISC machine and a central processing unit. Its primary features include a quad-core ARM Cortex A53 processor with a 64-bit CPU running at 1.2 GHz and 1 GB of RAM.

The device is the size of a credit card and has 4 poles stereo output, CSI camera port, composite video port, wireless LAN (802.11 b/g/n), and Bluetooth 4.1 chipset.

B) SOFTWARE BASED OBJECT DETECTION AND RECOGNITION MODULE (SODRM)

The system uses the MobileNet architecture for object detection and recognition, which is lightweight and provides accurate results, making it suitable for low computational devices like DSPs. A CNN is trained using TensorFlow API in Python programming language for object detection.

C) MOBILENET ARCHITECTURE

The Mobile-Net architecture is a lightweight deep neural network that uses depth-wise separable convolution layers. It consists of 21 layers, including both simple and deep convolutional layers. Simple convolution layers apply convolution using a 3x3 kernel, followed by batch normalization and ReLU activation. Deep layers have six stages, with the first three steps identical to the simple layers, followed by a 1x1 convolutional operation, batch normalization, and ReLU activation.

D) STRUCTURAL SIMILARITY INDEX (SSIM)

To save computational power and make the system real-time, it is better to check the similarity between two successive frames in a video signal. The SSIM index is used to measure this similarity, and if the index is less than a certain threshold T (which is set to 0.7 in this work), the current frame is processed. Otherwise, it is skipped. The SSIM index is calculated using the mean, variance, and covariance of the frames, along with stabilizing constants S1 and S2.

E) TEXT TO SPEECH (SAPI)

In this project, the Speech API (SAPI 5.3) [43] is utilized to generate audio based on the input text, which is obtained from the CNN. Specifically, SAPI takes the textual input that consists of the object names detected in the current frame and generates an audio signal that is sent to headphones. This enables visually impaired people to hear the names of the objects detected in the frame. The SAPI is a widely used software development kit that enables developers to create applications that can generate speech. It is a part of the Microsoft Windows operating system and is designed to help users with visual impairments interact with the system. The SAPI provides a set of tools and functions that can be used to generate speech from a wide range of sources, including text, XML, and other data formats. Overall, the integration of SAPI into this project enhances the accessibility of the system for visually impaired users by enabling them to receive audio feedback of the objects detected in the video stream.

F) ENCODER

In this system, after the objects are detected and recognized by the CNN, snapshots of the frames with bounding boxes around the objects and their corresponding labels are generated. These snapshots are then encoded using the JPEG encoder and stored in a database. The web interface allows easy access to these snapshots, making it possible for users to view the detected objects and their labels in real-time. This process facilitates quick and easy identification of the objects captured by the video feed.

G) GPS AND GSM MODULE

The proposed system [44] uses the Pi Anywhere 4G and LTE Hat for the Raspberry Pi Beta. This module is designed to provide 4G mobile data, GPS positioning information, and battery support, and can be easily plugged into the Raspberry Pi minicomputer. With ultra-fast 4G connectivity (100 Mbps down/ 50 Mbps up), it enables seamless video streaming and downloads. Additionally, the onboard GPS module allows for easy access to location data, making it ideal for tracking the user's location. The module is depicted in Figure 5, which showcases the GPS and GSM chips used in the module.

H) WEB-BASED APPLICATION

The web interface was developed to ensure the safety of VIPs by enabling family members to track their movements and snapshots from home. The web server running on a DSP periodically sends real-time GPS coordinates and labeled snapshots to the web interface, which maps the received coordinates onto a map using Google-map API. The interface displays the most recent labeled image and stores all received images in a gallery for review.

The web-based application was developed using Django, allowing users to log in and monitor visually impaired persons via a web browser. Upon logging in, users are directed to a dashboard displaying the most recent snapshot and real-time location of the VIP. The interface receives live video feeds and real-time coordinates, which are mapped onto a map using Google map API.

1V. BLOCK DIAGRAM

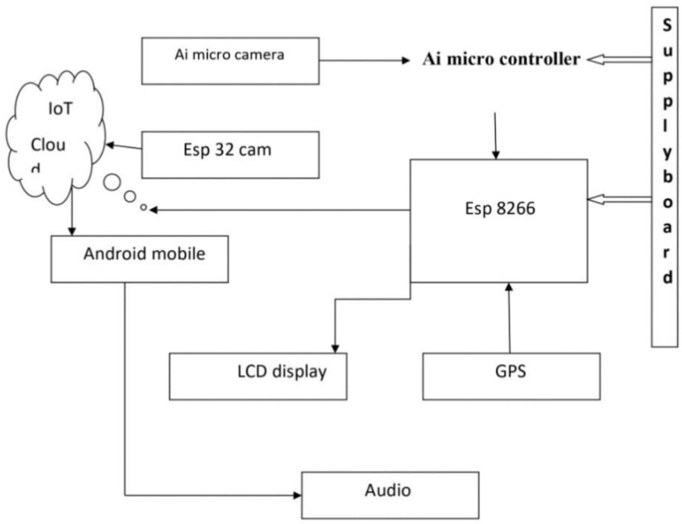

Fig 2 – Block diagram

The block diagram represents the operation of the entire system through a network of interconnected components. Each component plays a specific role in the overall functionality of the system, and technical specifications and information are provided in corresponding text boxes. Notably, the diagram includes a cloud drawing in the upper left corner, indicating the system's integration with IoT cloud-based services.

The central part of the diagram features five instances of "ESP 8256," suggesting the significance of this component to the system's functionality. This diagram provides valuable insight into how GPS technology functions and how the various components work together to facilitate accurate location tracking.

V. RELATED WORKS

A) R-CNN

Regional-based convolution neural networks (R-CNN) are a technique that extracts feature vectors from region proposals to detect objects in images. These feature vectors are then evaluated with Support Vector Machines (SVM) for each class. While R-CNN can achieve high accuracy, it is not able to operate in real-time due to the expensive training process and the inefficiency of region proposition, even with faster variants such as Fast R-CNN and Faster R-CNN.

B) YOLO

You Only Look Once (YOLO) [10] is a technique designed for one-step detection and classification of objects in an input image. It predicts bounding boxes and class probabilities in one evaluation of the image. YOLO's fastest architecture achieves up to 45 FPS, while a smaller version, Tiny-YOLO, can reach up to 244 FPS (Tiny YOLOv2) on a GPU-powered computer.

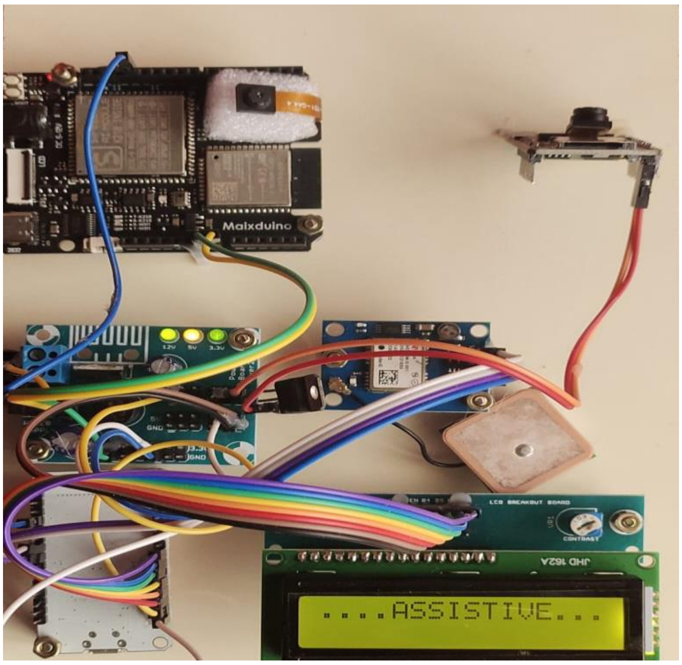

Fig 3 – Assistive system

Unlike traditional systems, YOLO performs bounding box and class predictions simultaneously. The input image is divided into an S x S grid, where B bounding boxes are defined in each grid cell, each with a confidence score. Confidence is the probability that an object exists in each bounding box and is defined as C = Pr(Object) * IOUtruth_pred (1), where IOU represents the intersection over union between the predicted bounding box and ground truth. YOLO's approach is unique in that it predicts bounding boxes and class probabilities together, resulting in efficient and accurate object detection.

VI. COMPARATIVE ANALYSIS

In this section, we present a quantitative and comparative analysis of the proposed system in comparison to existing assistive devices. Our evaluation is based on essential attributes such as concise and clear information delivery within a timely manner, reliable performance throughout the day and night, proper functioning in both indoor and outdoor environments, real-time analysis capabilities, and a high accuracy rate.

To assess each feature, we have assigned a weightage, with a device containing all features achieving a weightage of 10. Devices that operate only in daytime or indoor environments are given a score of 5. The evaluated features are fundamental to assistive device design, and the highest weightage is assigned to a device that possesses all features. The value of a device is represented as Vk, ranging from 0-10, with a score of 10 awarded to a device containing all features. Only devices that have undergone real-time scenario testing are evaluated, and user feedback is also taken into account.
In summary, we have conducted a comprehensive analysis of the proposed system and compared it to existing assistive devices based on critical attributes. Our evaluation methodology considers multiple factors and assigns appropriate weightages to each feature, ensuring a fair and objective comparison.

Date	Lattitude	Longitude					
2/28/2023	13.032542	80.209948	0.11	4	-2.7	7093500	280223
2/28/2023	13.032542	80.209949	0.85	4	-2.3	7094400	280223
2/28/2023	13.032557	80.209954	0.06	4	-0.8	7095300	280223
2/28/2023	13.032569	80.209952	0.11	4	-1.9	7100100	280223
2/28/2023	13.032578	80.209946	0.06	4	-3.6	7101000	280223
4/1/2023	13.032544	80.209991	4.39	4	20.2	12433400	10423
4/1/2023	13.032566	80.209938	2.83	4	21.8	12434100	10423
4/1/2023	13.032582	80.209905	1.52	4	24.2	12434900	10423
4/1/2023	13.032595	80.209892	2.35	4	24.5	12453400	10423
4/1/2023	13.032582	80.209919	7.45	4	27.2	12454100	10423
4/1/2023	13.032574	80.209932	3.74	4	24.2	12454700	10423
4/1/2023	13.032571	80.209932	4.32	4	23.5	12455400	10423
4/1/2023	13.032572	80.209933	3.57	4	25	12460000	10423
4/1/2023	13.032576	80.209932	2.76	5	27	12460700	10423
4/1/2023	13.032579	80.20994	0.28	5	29.1	12461300	10423
4/1/2023	13.032569	80.209982	0.15	5	28.9	12462000	10423
4/1/2023	13.032565	80.209993	0.06	5	28.9	12462700	10423
4/1/2023	13.03256	80.209995	0.04	5	27.7	12463300	10423
4/1/2023	13.032561	80.209982	0.11	7	26.6	12464000	10423
4/1/2023	13.032566	80.209979	0.04	7	28.8	12464700	10423
4/1/2023	13.032571	80.209979	0	7	31.5	12465400	10423
4/1/2023	13.03257	80.209977	0.02	7	31.8	12470100	10423
4/1/2023	13.032568	80.209977	0.04	8	30.5	12470700	10423
4/1/2023	13.032567	80.209977	0.04	8	29.6	12471400	10423
4/1/2023	13.032565	80.209975	0.09	8	28.2	12472100	10423
4/1/2023	13.032563	80.209973	0.22	7	27.5	12472800	10423
4/1/2023	13.032564	80.20997	0.04	8	28.4	12473600	10423
4/1/2023	13.032565	80.20997	0.04	8	29.6	12474300	10423
4/1/2023	13.032566	80.20997	0	8	30.3	12474900	10423
4/1/2023	13.032567	80.209969	0.04	8	30.1	12475600	10423
4/1/2023	13.032562	80.209962	0.04	8	26.9	12480300	10423
4/1/2023	13.032558	80.209958	0.13	8	25	12481000	10423
4/1/2023	13.032557	80.20996	0.11	8	24.6	12481800	10423
4/1/2023	13.032559	80.209965	0.07	8	26	12482400	10423
4/1/2023	13.03256	80.209963	0.04	8	26.8	12483100	10423
4/1/2023	13.032557	80.209963	0.09	8	26.5	12483900	10423
4/1/2023	13.032551	80.209966	0.17	8	24.6	12484300	10423
4/1/2023	13.032549	80.209972	0.15	8	23.1	12484900	10423
4/1/2023	13.032548	80.209976	0.33	8	22.3	12485500	10423
4/1/2023	13.032547	80.209983	0.04	9	21.8	12490200	10423
4/1/2023	13.032549	80.209985	0.15	9	22	12490900	10423
4/1/2023	13.032554	80.20998	0.11	9	23.2	12491700	10423
4/1/2023	13.032556	80.209981	0.04	9	24	12492300	10423
4/1/2023	13.032558	80.20998	0.26	9	24.5	12493100	10423
4/1/2023	13.032559	80.209978	0.15	9	24.6	12493700	10423
4/1/2023	13.032559	80.209977	0.13	9	24.3	12494400	10423
4/1/2023	13.032561	80.209976	0.09	9	24.1	12495100	10423
4/1/2023	13.032563	80.209973	0	9	24.2	12495800	10423
4/1/2023	13.032568	80.209973	0.41	9	25.3	12500500	10423
4/1/2023	13.032568	80.209971	0.28	9	25	12501200	10423

Fig 4 – Comparative analysis

VII. COMPARISON WITH OTHER FAST OBJECT DETECTION NETWORKS

Currently, there is a dearth of real-time object detection algorithms available for non-GPU devices. Despite being one of the fastest options available, YOLO-LITE's processing speed of approximately 2.4 FPS is still insufficient for non-GPU computers. In contrast, Google's object detection API provides several lightweight architectures in its model zoo, among which SSD Mobilenet V1 stands out as particularly impressive.

One of the techniques employed by MobileNet to achieve its high performance is the use of depthwise separable convolutions. These convolutions combine depthwise convolution and pointwise convolution to reduce the model's size while retaining the same amount of learned information in each convolution. This approach may partially explain the higher mAP results obtained with SSD MobileNet COCO.

To summarize, while YOLO-LITE remains a top choice for real-time object detection algorithms, its processing speed may still fall short for non-GPU devices. Google's object detection API offers a number of lightweight architecture options, among which SSD Mobilenet V1 stands out as particularly impressive, thanks in part to its use of depthwise separable convolutions.

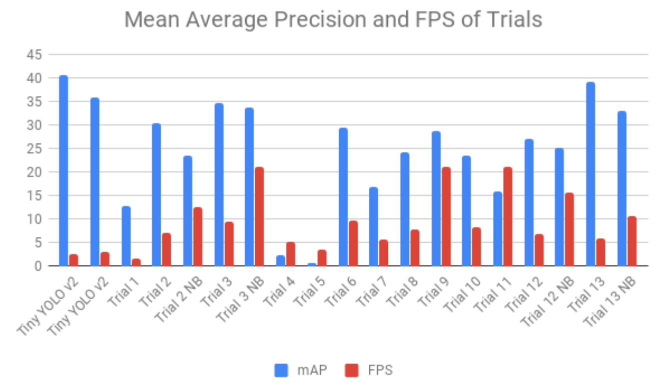

Fig 5- Mean Average Precision and FPS of Trials

VIII. FUTURE WORKS

In general, lightweight object detection models like YOLO-LITE trade off accuracy for faster processing speed. While YOLO-LITE is currently the fastest state-of-the-art model, its low mAP limits its applicability in real-world scenarios like autonomous vehicles. To improve accuracy, pre-training the network on ImageNet or using R-CNN to find and classify bounding boxes separately could be effective techniques. Another potential approach is to implement multiple prediction locations or filter pruning to make the network more efficient and compact. Finally, ShuffleNet's group convolution and channel shuffling could be used to reduce computation while maintaining information during training.

IX. CONCLUSION

In this paper, a system is proposed to assist VIPs in mobility and ensure their safety. The system is designed based on the daily requirements of VIPs, which includes object recognition and sensing the natural environment using a CNN-based low-power Mobile-Net architecture. A web-based application is also developed to enable the VIP's family to monitor their location and movement using live feed from the camera.

The experimental analysis conducted shows that the proposed system outperforms other devices in terms of supported features. There were 18 different trials conducted during the experimentation phase, and the mAP and FPS for each trial and Tiny-YOLOv2 were recorded. While the development for YOLO-LITE was done on PASCAL VOC, the best trial was also run on COCO.

This paper proposes a system for VIPs to assist them in mobility and ensure their safety, using a low-power Mobile-Net architecture to recognize objects and a web-based application for location sharing. The proposed system outperformed other devices in terms of supported features. In the future, aesthetic sense and bone conduction headphones may be considered to evaluate performance. The YOLO-LITE architecture was successfully trained for both VOC and COCO and implemented as a web-based model. The FPS may differ depending on the device.

X. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the organizers of this conference for providing us with the opportunity to present our research. We also extend our thanks to the reviewers for their insightful feedback, which has helped us improve the quality of our paper.

Furthermore, we would like to acknowledge the contributions of our research team, who have worked diligently to gather and analyze the data, and develop the proposed system. We also wish to thank peers for their valuable inputs and suggestions, which have helped us refine our ideas and methodology.

XI. REFERENCES

[1] F. Ashiq et al., *"CNN-Based Object Recognition and Tracking System to Assist Visually Impaired People,"* in IEEE Access, vol. 10, pp. 14819 -14834, 2022, doi: 10.1109/ACCESS.2022.3148036.

[2] Work Sheet. Accessed: Mar. 7, 2020. [Online]. Available: https://www.who.int/en/newsroom/factsheets/detail/blindness-and visual -impairment

[3] R. Velázquez, *''Wearable assistive devices for the blind,''* in Wearable and Autonomous Biomedical Devices and Systems for Smart Environment. Berlin, Germany: Springer, 2010.

[4] L. B. Neto, F. Grijalva, V. R. M. L. Maike, L. C. Martini, D. Florencio, M. C. C. Baranauskas, A. Rocha, and S. Goldenstein, *''A Kinect-based wearable face recognition system to aid visually impaired users,''* IEEE Trans. Human-Mach. Syst., vol. 47, no. 1, pp. 52–64, Feb. 2017.

[5] C. Shah, M. Bouzit, and L. Vasquez, *''Evaluation of RU-netratactile feedback navigation system for the visually impaired,''* in Proc. Int. Workshop Virtual Rehabil., 2006, pp. 72–77.

[6] M. R. U. Saputra, Widyawan, and P. I. Santosa, *''Obstacle avoidance for visually impaired using auto-adaptive thresholding on Kinect's depth image,''* in Proc. IEEE 11th Int. Conf. Ubiquitous Intell. Comput., IEEE 11th Int. Conf. Auton. Trusted Comput., IEEE 14th Int. Conf. Scalable Comput. Commun. Associated Workshops, Dec. 2014, pp. 337–342.

[7] Y. Yi and L. Dong, *''A design of blind-guide crutch based on multisensors,''* in Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discovery (FSKD), Aug. 2015, pp. 2288–2292.

[8] K. Kumar, B. Champaty, K. Uvanesh, R. Chachan, K. Pal, and A. Anis, *''Development of an ultrasonic cane as a navigation aid for the blind people,''* in Proc. Int. Conf. Control, Instrum., Commun. Comput. Technol. (ICCICCT), Jul. 2014, pp. 475–479.

[9] C. T. Patel, V. J. Mistry, L. S. Desai, and Y. K. Meghrajani, *''Multisensor– based object detection in indoor environment for visually impaired people,''* in Proc. 2nd Int. Conf. Intell. Comput. Control Syst. (ICICCS), Jun. 2018, pp. 1–4.

[10] Z. Bauer, A. Dominguez, E. Cruz, F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, *''Enhancing perception for the visually impaired with deep learning techniques and low-cost wearable sensors,''* Pattern Recognit. Lett., vol. 137, pp. 27–36, Sep. 2020.

[11] L.-B. Chen, J.-P. Su, M.-C. Chen, W.-J. Chang, C.-H. Yang, and C. -Y. Sie, *''An implementation of an intelligent assistance system for visually impaired/blind people,''* in Proc. IEEE Int. Conf. Consum. Elect ron. (ICCE), Jan. 2019, pp. 1–2.

[12] M. Poggi and S. Mattoccia, *''A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning,''* in Proc. IEEE Symp. Comput. Commun. (ISCC), Jun. 2016, pp. 208–213.

[13] S.-H.Chae, *''Collision detection method using image segmentation for the visually impaired,''* IEEE Trans. Consum. Electron., vol. 63, no. 4, pp. 392–400, Nov. 2017.

[14] P. Salavati and H. M. Mohammadi, *''Obstacle detection using GoogleNet,''* in Proc. 8th Int. Conf. Comput. Knowl. Eng. (ICCKE), Oct. 2018, pp. 326–332.

[15] T. H. Nguyen, T. L. Le, T. T. H. Tran, N. Vuillerme, and T. P. Vuong, *''Antenna design for tongue electrotactile assistive device for the blind and visually-impaired,''* in Proc. 7th Eur. Conf. Antennas Propag. (EuCAP), 2013, pp. 1183–1186.

[16] J. Xiao, K. Ramdath, M. Iosilevish, D. Sigh, and A. Tsakas, *''A low cost outdoor assistive navigation system for blind people,''* in Proc. IEEE 8th Conf. Ind. Electron. Appl. (ICIEA), Jun. 2013, pp. 828–833.

[17] S. Bharambe, R. Thakker, H. Patil, and K. M. Bhurchandi, *''Substitute eyes for blind with navigator using android,''* in Proc. Texas Instrum. India Educ. Conf., Apr. 2013, pp. 38–43.

[18] A. S. Martinez-Sala, F. Losilla, J. C. Sánchez-Aarnoutse, and J. García-Haro, ''Design, implementation and evaluation of an indoor navigation system for visually impaired people,'' Sensors, vol. 15, pp. 32168–32187, Dec. 2015.

[19] Aladrén, G. López-Nicolás, L. Puig, and J. J. Guerrero, *''Navigation assistance for the visually impaired using RGB-D sensor with range expansion,''* IEEE Syst. J., vol. 10, no. 3, pp. 922–932, Sep. 2016.

[20] A.Yamashita, K. Sato, S. Sato, and K. Matsubayashi, *''Pedestrian navigation system for visually impaired people using HoloLens and RFID,''* in Proc. Conf. Technol. Appl. Artif. Intell. (TAAI), Dec. 2017 , pp. 130–135.

[21] A.Mancini, E. Frontoni, and P. Zingaretti, *''Mechatronic system to help visually impaired users during walking and running,''* IEEE Trans. Intell. Transp. Syst., vol. 19, no. 2, pp. 649–660, Feb. 2018. J. Schmidhuber, "Deep learning in neural networks: An overview, Neural networks, vol. 61, pp. 85–117, 2015. 1

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik, *"Rich feature hierarchies for accurate object detection and semantic segmentation,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 580–587. 1

[23] C. Szegedy, A. Toshev, and D. Erhan, *"Deep neural networks for object detection,"* in Advances in neural information processing systems, 2013, pp. 2553–2561. 1

[24] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, J. Kindelsberger, L. Ding, S. Seaman et al., *"Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation,"*arXiv preprint arXiv:1711.06976, 2017. 1

[25] O. Akgul, H. I. Penekli, and Y. Genc, *"Applying deep learning in augmented reality tracking,"* in Signal-Image Technology &InternetBased Systems (SITIS), 2016 12th International Conference on. IEEE, 2016, pp. 47–54. 1

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *"Rethinking the inception architecture for computer vision,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826. 1 [27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, *"Inception-v4, inception-resnet and the impact of residual connections on learning."* in AAAI, vol. 4, 2017, p. 12. 1

[27] K. He, X. Zhang, S. Ren, and J. Sun, *"Deep residual learning for image recognition,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778. 1

[28] K. Simonyan and A. Zisserman, *"Very deep convolutional networks for large-scale image recognition,"*arXiv preprint arXiv:1409.1556, 2014. 1

[29] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, *"You only look once: Unified, real-time object detection,"* in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788. 1, 2, 6, 7 [11] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, *"Rich feature hierarchies for accurate object detection and semantic segmentation,"*CoRR, vol. abs/1311.2524, 2013. [Online]. Available: http://arxiv.org/abs/1311.2524 1, 2 [31] R. B. Girshick, *"Fast R-CNN,"*CoRR, vol. abs/1504.08083, 2015. [Online]. Available: http://arxiv.org/abs/1504.08083 2

[30] W. Elmannai and K. Elleithy, *''Sensor-based assistive devices for visually impaired people: Current status, challenges, and future directions,''* Sen sors, vol. 17, no. 3, p. 565, 2017.