IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

VOICE GUIDANCE BASED OBJECT DETECTION USING YOLOV7 MODEL

[1]Mrs. J Ranganayaki [2]K. Sadu Sundar [3]S. Logesh [4]S Logunath Bose [5]N.S Lokesh Kumar

[1] Assistant Professor, Department of Computer Science and Engineering,
Bharath Institute of Higher Education and Research, Chennai, India- 600073.
[2,3,4,5]Students, Department of Computer Science and Engineering,
Bharath Institute of Higher Education and Research, Chennai, India- 600073.

Abstract-

The challenges faced by visually impaired individuals in detecting objects and animals are significant. Our proposed solution involves integrating the YOLOv7 algorithm with deep neural networks (DNN), leveraging the COCO dataset for object recognition. This system communicates detected objects through voice feedback, aiding visually impaired users in understanding their surroundings. The utilization of YOLOv7, a proven system recognized for its accuracy in object detection, can greatly improve the effectiveness of our solution. By harnessing deep learning algorithms, the system has the potential to deliver more precise and reliable results, thereby enhancing the user experience for visually impaired individuals. To ensure success, it's crucial that the system's user interface and audio feedback are designed to be intuitive and easily understandable for visually impaired users. Incorporating user testing and feedback from individuals with visual impairments will be essential in refining the system to meet their specific needs and preferences. Additionally, continuous updates and enhancements to the algorithm and dataset will be vital to maintaining and improving the system's accuracy and performance over time, ensuring its ongoing effectiveness in assisting visually impaired individuals.

Keywords-- YOLOv7 algorithm, Deep neural networks (DNN), COCO dataset, Object recognition, Voice feedback, Visually impaired individuals, User interface design, Accessibility, Deep learning algorithms

I. INTRODUCTION

Visual need is a state of misplaced the visual affirmation due to physiological or neurological variables. The halfway visual incapacity talks to the require of integration in the progression of the optic nerve or visual center of the eye, and incorporate up to visual impedance is the full nonattendance of the visual light affirmation. In this work, a principal, cheap, inviting client, sharp daze course framework is orchestrated and executed to development the compactness of both flabbergast and clearly disabled individuals in a particular region. The proposed work solidifies a wearable gear comprises of light weight daze take

after and sensor based hindrance region circuit is made to offer offer assistance the flabbergast individual to examine alone securely and to dodge any obstacles that may be experienced, whether settled or flexible, to evade any conceivable mischance.

There are number of flabbergast individuals in the society, who are driving forward in spite of the fact

that working out the vital things of day by day life and that may put lives at chance whereas voyaging. There is a require these days to permit security and security to stupor individuals. There have been few contraptions orchestrated so distant off to offer offer assistance the flabbergast. Visual lacking or visual failure is a condition that impacts different individuals around the world. The utilization of the flabbergast course framework is remarkably less and is not able. The daze traveler is subordinate on other arrange like white cane, data given by the individuals, orchestrated dogs etc. Different for all bury and purposes disabled individuals utilize strolling sticks or facilitate pooches to move from put to put. A facilitate pooch is organized for planning its clients to maintain a strategic distance from the mischances from objects and boundaries over a settled way or in a settled run. When a clearly impeded individual businesses a strolling take after, he waves his take after and finds the obstacle by striking the impediments in his way. The consider of as of presently made frameworks and examination of the execution methods utilized, driven us to characterize a unused framework which show up overcome the preventions in the past frameworks. Along these lines utilizing the existing advances we give a course of activity to the communicated issue. One more application is outlined out for family individuals to get to the daze person's zone through the server, at anything point needed.

This paper presents a framework concept to permit a sharp electronic offer help for flabbergast individuals. We propose to orchestrate an brilliantly contraption which cautions the individual on event of obstacles based on remove between the individual and the deterrent. Here, this brilliantly contraption not as it were cautions but in expansion takes after the extend of the individual and educates the current position of the individual to his relatives through the utilize of server.

II. LITERATURE SURVEY

One such contribution comes from Rahman and Sadi (2021), who introduce an IoT-enabled automated object recognition system designed to assist visually impaired individuals in both indoor and outdoor settings. Leveraging laser sensors and the Single Shot Detector (SSD) model with MobileNet and Tensorflow-lite, their system provides real-time identification of objects and currency notes. Feedback from 375 participants indicates positive responses regarding the system's utility, cost-effectiveness, and overall effectiveness in aiding object recognition tasks.

Similarly, Chang and Chen (2021) propose a wearable assistive system focused on improving pedestrian safety at zebra crossings. Their solution includes smart sunglasses, a waist-mounted intelligent device, and an intelligent walking cane, all equipped with deep learning algorithms for real-time zebra crossing recognition. Experimental results demonstrate promising accuracy in identifying zebra crossings, thereby providing timely alerts and information to visually impaired pedestrians to navigate urban environments safely.

Another notable contribution is presented by Rahman and Islam (2020), who develop a wearable electronic device aimed at facilitating independent travel for visually impaired individuals while ensuring real-time safety monitoring. Utilizing ultrasonic sensors, a PIR motion sensor, and an accelerometer, their system detects obstacles, ground irregularities, and sudden falls. Through a smartphone application, users receive audible instructions for navigation, while guardians are promptly notified in emergencies, thereby enhancing user safety and confidence.

Additionally, Bauer and Dominguez (2020) propose a system to enhance perception for visually impaired individuals using deep learning techniques and low-cost wearable sensors. Their approach delivers information about potential obstacles in outdoor scenarios through spoken or haptic feedback, achieving high accuracy in obstacle presence detection. Pilot tests involving individuals with vision impairments validate the system's effectiveness in improving navigation capabilities.

Finally, Tapu and Mocanu (2020) provide a comprehensive survey of wearable assistive devices for visually impaired individuals. Their work evaluates various systems based on features and performance parameters established with input from the blind community, offering valuable insights into the capabilities and limitations of existing assistive technologies. This survey informs future research directions and underscores the importance of ongoing innovation in this field to enhance the quality of life for visually impaired individuals.

III. METHODOLOGY

To establish voice guidance-based object detection using the YOLOv7 model, the methodology involves several key steps. Initially, we define the specific objectives of the project, such as identifying the types of objects to be detected and the environments where the system will operate. Following this, we collect a diverse dataset of images or videos containing relevant objects and meticulously annotate them with bounding boxes and class labels to serve as training data. The YOLOv7 model is then chosen as the foundation for object detection due to its balance between accuracy and speed. We may also consider customizing the model to better suit the requirements of our application. Preprocessing steps are undertaken to prepare the dataset for training, including resizing, normalization, and augmentation.

Subsequently, we train the YOLOv7 model on the annotated dataset, fine-tuning its parameters as needed. Throughout the training process, we monitor key metrics such as loss and mean average precision (mAP) to gauge the model's performance and make necessary adjustments. Validation procedures are conducted using separate validation sets, and real-world testing is performed to ensure the model's effectiveness in practical scenarios. Feedback from visually impaired individuals is solicited and incorporated to enhance the system's usability and accessibility.

In parallel, we develop a voice guidance system that works in tandem with the YOLOv7 model to provide real-time audio feedback on detected objects. This system is designed to deliver clear and concise instructions or descriptions of identified objects to aid users in navigation and situational awareness. The integration of the voice guidance system with the object detection pipeline is carefully implemented to ensure synchronization and seamless operation.

Finally, rigorous testing is conducted to evaluate the integrated system's performance and robustness. Any identified issues are addressed, and the system is deployed in real-world settings, with ongoing maintenance and updates to ensure its continued effectiveness. Through this iterative process, we aim to create a reliable and user-friendly solution that enhances the mobility and safety of visually impaired individuals in various environments.

EXISTING SYSTEM

The existing approach for counting objects in camera-captured images focuses on amalgamating intra-camera visual features and exploring inter-camera knowledge to improve accuracy and scalability. However, there are opportunities to refine this solution further. Firstly, advanced feature fusion techniques could be integrated to better combine intra-camera visual features. This might involve utilizing sophisticated deep learning architectures such as Siamese networks or attention mechanisms to extract and merge pertinent features from various camera perspectives, thereby bolstering the discrimination capabilities of the counting model.

In addition to feature fusion, incorporating contextual information could significantly enhance counting precision. By integrating scene semantics or historical data, the system can gain a deeper understanding of the context in which objects are detected, thereby leading to more accurate counting outcomes. Moreover, the development of dynamic crowd density estimation techniques could allow for adaptive adjustments in counting strategies based on real-time crowd density analysis. This dynamic approach ensures accurate counting in scenarios with fluctuating congestion levels, enhancing the system's resilience across diverse environmental conditions.

Addressing variations in camera perspectives and viewpoints through cross-view calibration techniques is another critical aspect of refining the existing solution. By calibrating camera parameters and aligning different camera views, the system can more accurately match objects across views, minimizing counting discrepancies caused by perspective distortions. Additionally, improving the robustness of blob matching algorithms is essential to handle complex scenarios such as occlusions or partial visibility. Advanced matching criteria and outlier rejection mechanisms can enhance the accuracy of entity alignment across camera views, resulting in more reliable counting outcomes.

Introducing real-time feedback mechanisms to monitor counting performance and provide adaptive adjustments is also crucial. By continuously analyzing counting results and user feedback, the system can dynamically fine-tune counting strategies to optimize accuracy and scalability. Furthermore,

exploring the integration of multi-sensor data, such as depth sensors or infrared cameras, could complement visual information and enhance object detection and tracking capabilities. Leveraging additional sensor modalities improves counting accuracy, particularly in challenging lighting conditions or low-visibility scenarios.

Existing System Disadvantages:

The current methodology for counting objects in camera-captured images lacks effectiveness in utilizing intra-camera visual features, leveraging heterogeneous information, and ensuring successful knowledge sharing across views.

PROPOSED SYSTEM

In the proposed system, multi-view object detection using YOLO V7 in real-time is emphasized as a critical task with broad applicability across various domains within computer vision. These applications span diverse fields such as video analytics, robotics, autonomous vehicles, multi-object tracking, object counting, and medical image analysis. An object detector serves as a foundational component in these applications, enabling tasks like object localization and classification.

Operating on the principle of convolutional neural networks (CNNs), object detectors like YOLO V7 extract features from input images to predict bounding boxes and class probabilities for identified objects. This methodology offers robustness and efficiency in real-time object detection tasks, ensuring timely and accurate recognition across different contexts and scenarios.

The real-time aspect of object detection is particularly crucial in dynamic environments where objects may move or change rapidly. YOLO V7's capability to achieve high-speed inference while maintaining accuracy is vital for applications requiring instantaneous responses, such as robotics, surveillance, and autonomous vehicles.

By integrating multi-view object detection using YOLO V7, the proposed system aims to provide a comprehensive solution for various computer vision challenges. Its ability to handle real-time processing and accurately detect objects across multiple views makes it well-suited for a wide range of applications, ultimately contributing to advancements in fields reliant on computer vision technology.

proposed system advantages:

YOLOv7 is widely acclaimed for its unparalleled speed and precision, making it the preferred choice for real-time object detection tasks across various computer vision applications.

Within this project, methodologies are broadly categorized into two main streams: machine learning-based and deep learning-based approaches.

In machine learning-based approaches, the initial step involves defining features, often utilizing techniques like YOLOv7 for subsequent classification tasks.

Conversely, deep learning techniques, particularly those anchored in convolutional neural networks (CNNs), offer the advantage of end-to-end object detection, negating the necessity for explicit feature definition.

Deep learning models, including YOLOv7, excel in capturing intricate patterns and features directly from raw data, resulting in more robust and accurate object detection without relying on predefined features.

IV. SYSTEM IMPLEMENTATION

YOLO V7 FRAMEWORK

Object revelation is a computer vision errand that joins both localizing one or more objects insides an picture and classifying each address in the picture. It is a challenging computer vision errand that requires both compelling address localization in organize to find and draw a bounding box around each address in an picture, and address classification to foresee the change lesson of address that was localized.

CLASS Title FORMATION

COCO Method

Though the layers are colloquially suggested to as convolutions, this is as it were by tradition. Numerically, it is in reality a sliding bit thing or cross-correlation. This has centrality for the records in the organize, in that it impacts how weight is chosen at a particular record point.

IMAGE GENERATOR MODULE

METHOD: CONVOLUTIONAL Organize LAYERS

Convolutional systems may connect neighborhood or around the world pooling layers to streamline the basic computation. Pooling layers reduce the estimations of the information by combining the yields of neuron clusters at one layer into a single neuron in the another layer. Adjoining pooling combines little clusters, commonly 2 x 2. Around the world pooling acts on all the neurons of the convolutional layer. In expansion, pooling may compute a max or an standard. Max pooling businesses the most uncommon respect from each of a cluster of neurons at the earlier layer. Conventional pooling employments the standard respect from each of a cluster of neurons at the earlier layer.

CLASSIFIERS MODULE

METHOD: YOLO Course of activity AND WEIGHTS

YOLO V7 consider classification as one of the most energetic investigate and application zones. Yolo V7 is the division of Fake Encounters (AI). The neural organize was orchestrated by Yolo V7 calculation. The specific combinations of capacities and its influence whereas utilizing Yolo V7 as a classifier is reviewed and the rightness of these capacities are analyzed for particular sorts of datasets. The Yolo V7 can be utilized as a altogether profitable gadget for dataset classification with fitting combination of arranging, learning and exchange capacities. When the most uncommon probability strategy was compared with COCO method, the Yolo V7 was more correct than most uncommon probability technique. A tall prescient capacity with tenacious and well working Yolo V7 is conceivable. It outlines to be more productive than other classification algorithms.

The approach consolidates a single noteworthy convolutional neural organize (at to begin with a alteration of GoogLeNet, a whereas afterward overhauled and called DarkNet based on VGG) that parts the input into a organize of cells and each cell clearly predicts a bounding box and address classification. The result is a wide number of candidate bounding boxes that are set into a last crave by a post-processing step.

There are three basic combinations of the approach, at the time of composing; they are YOLOv1, YOLOv2, and YOLOV7. The to begin with outline proposed the common building, while the miniature outline refined the organize and made utilize of predefined stay boxes to move forward bounding box suggestion, and outline three advance refined the outline building and arranging process.

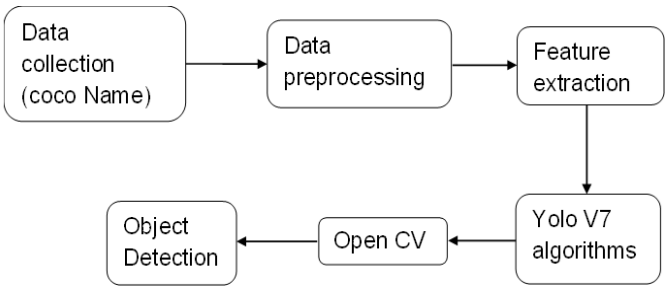
CREATE AND SAVA Show

These were organized utilizing the DarkNet code base on the MSCOCO dataset. Download the show up weights and put them into your current working registry with the filename “yoloV7.weights.” It is a clearing record and may take a scaled down to download depending on the speed of your web connection.

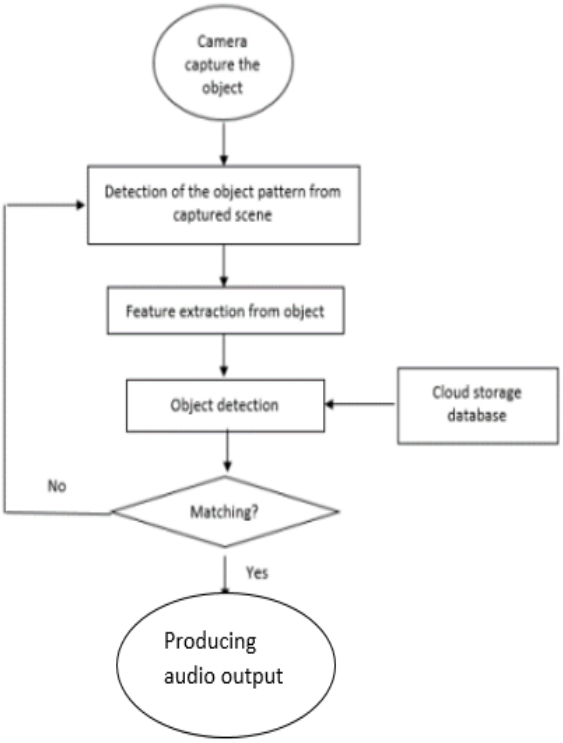
Next, we require to characterize a Keras outline that has the right number and sort of layers to orchestrate the downloaded show up weights. The show up arrange is called a “DarkNet” and was at to begin with straightforwardly based on the VGG-16 model.

Next, we require to stack the show up weights. The show up weights are put truant in anything organize that was utilized by DarkNet. Or maybe than endeavoring to disentangle the record physically, we can utilize the WeightReader course given in the script. To utilize the WeightReader, it is instantiated with the way to our weights record (e.g. ‘yoloV7.weights‘). This will parse the record and stack the show up weights into memory in a organize that we can set into our Keras show up and expect the challenge will be detected.

V. System Architecture



FLOWCHART:



VI. Result and Discussion

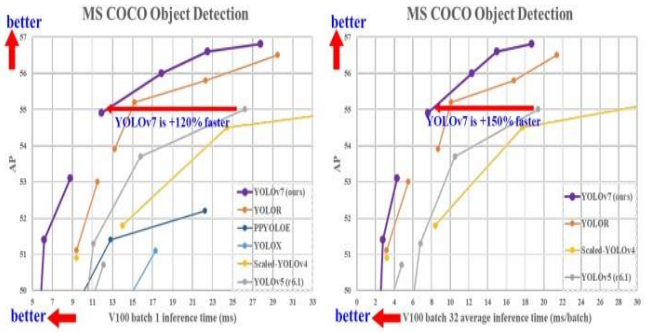


Figure 1

Figure 1. Different between YOLOV3 and YOLOV7 algoithms using Graphs

These graphs illustrate the performance trade-offs between different object detection models in terms of speed and accuracy. YOLOv7 stands out as significantly faster than other models like YOLOV3 while maintaining high accuracy. The percentage difference in speed gain for YOLOV7 between the two graphs indicates varying performance based on processing single images or multiple images together.

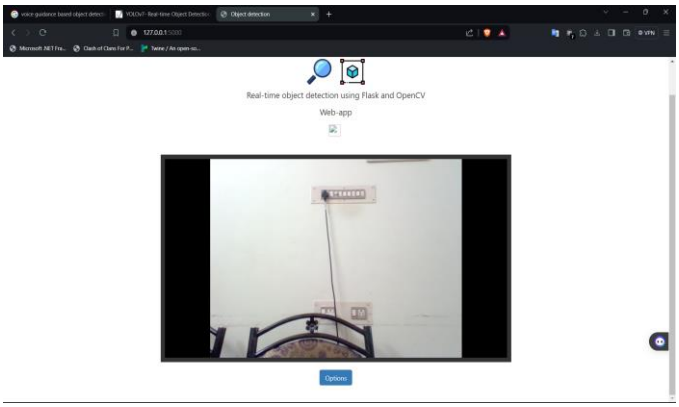


Figure 2: The above image is the algorithm before detection

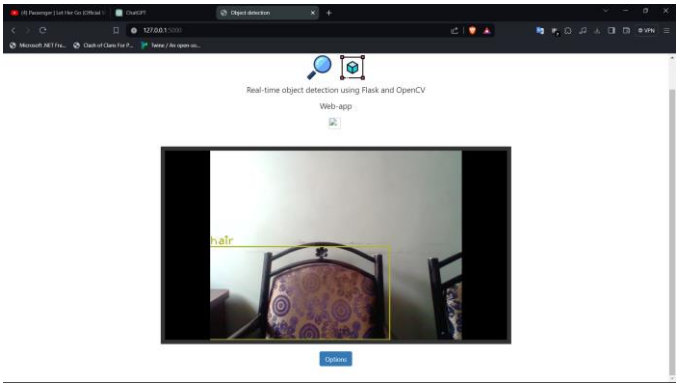


Figure 3: The above image detects a object by bounding box with text

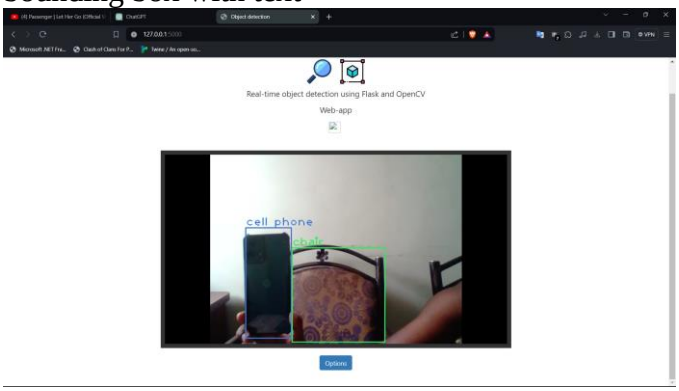


Figure 3: The above image detects multiple objects in the frame.

VII. Conclusion

This paper displayed a keen and shrewdly framework for VIPs to help them in portability and guarantee their security. The proposed framework is based on the day-to-day necessities of VIPs. It helps them in visualizing the environment and giving a sense of the environment. They can recognize objects around them and sense the characteristic environment utilizing DNN-based low-power Mobile-Net engineering. Additionally, a web-based application is created to guarantee the security of VIPs. The

client of this application can turn the on-demand work to share his/her area with the family. It is useful to their family as they can screen the development of VIPs and can track his/her area utilizing the live nourish from the camera. The exploratory examination appears that the proposed framework given palatable comes about and outflanked other gadgets in terms of backed highlights. The errand is exceptionally challenging but utilizing state-of-the-art strategies, the appropriateness of a gadget can be computed.

VIII. REFERENCES

[1] W. Elmannai and K. Elleithy, ‘‘Sensor-based assistive devices for visuallyimpaired people: Current status, challenges, and future directions,’’ Sensors, vol. 17, no. 3, p. 565, 2017.

[2] Work Sheet. Accessed: Mar. 7, 2020. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/blindness-andvisual-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, M. Youssef, 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. Electron. (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, M.-C. Kang, J.-Y. Sun, B.-S. Kim, and S.-J. Ko, ‘‘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] A. 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.