Robotic Guide Dog for Real-time Indoor Object Detection and Classification with Localization

Nathan Rees
UTS Robotics Institute
University of Technology Sydney
Sydney, Australia.
Nathan.Rees@uts.edu.au

Karthick Thiyagarajan
UTS Robotics Institute
University of Technology Sydney
Sydney, Australia.
Karthick.Thiyagarajan@uts.edu.au

Sarath Kodagoda
UTS Robotics Institute
University of Technology Sydney
Sydney, Australia.
Sarath.Kodagoda@uts.edu.au

2024 IEEE Applied Sensing Conference (APSCON) | 979-8-3503-1727-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/APSCON60364.2024.10465953

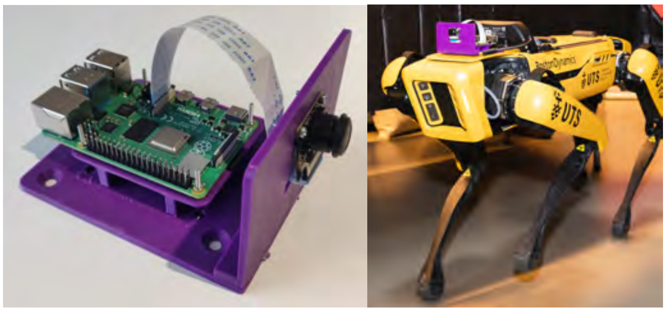
***Abstract*—Guide dog robots with advanced sensing abilities could be a big boon to vision-impaired people as some of them may choose technological solutions over real-life guide dogs. In this study, we propose a method that combines a robotic guide dog sensing system with the YOLO-GUIDE framework to enable real-time indoor object detection and classification with localization. The performance was assessed using ten indoor objects. The qualitative test outcomes showed the effectiveness of the proposed method, while quantitative evaluation results with 0.76 Precision, 0.67 Recall, and a 0.71 F1-score indicate high performance. The YOLO-GUIDE proved its superiority by outperforming other relevant models.**

***Index Terms*—Sensor applications, blind, assistive robotics, YOLO, object detection and classification, object localization.**

I. Introduction

The World Health Organization estimates that there are at least 285 million visually impaired persons (VIPs) in the globe [1]. The VIPs often employ a variety of mobility aids in their daily mobility activities, including white canes, hand-held smart devices [2], blind sticks with ultrasonic sensors [3], electromagnetic sensors [4], and electronic mobility canes [5]. Guide Dogs, on the other hand, are commonly recognized as a crucial mobility aid by VIPs since they facilitate more fluid movement for the user than other available aids. Even though they greatly increase the degree of freedom for VIPs, raising and training guide dogs is costly and laborious. Additionally, trained dogs have unique skill sets, exhibit distinct behaviors in various contexts, and cannot learn from other trained dogs. Technological solutions may solve some of such limitations.

Guide Dogs NSW/ACT is Australia's leading provider of guide dogs. In partnership with Guide Dogs NSW/ACT, the authors have investigated the capabilities of both mobility aids and advanced robotic technologies against a set of fundamental functional mobility aid features with the purpose of developing a minimum viable product to support VIPs in daily mobility tasks [6]. The findings reveal that technology advancements are restricted in scope, concentrating on just a few of the identified needed elements, and do not offer the capabilities of a fully working robot guide on their own. The detection and classification of objects in their surroundings are arguably one of the most important of the seven traits specified by the study for robotic guide dogs to increase VIPs independence and confidence while exploring the world.



(a) (b)

Fig. 1: (a) Sensing module, and (b) Guide dog robot integrated with the sensing module for obtaining real-time visual data.

According to the Facebook research study [7], YOLO models outperform top detection techniques like DPM and R-CNN by a wide margin and produce less than half as many background errors as Fast R-CNN. The YOLO-based models and their architectural successors set themselves apart from other relevant models like Mask R-CNN [8], Faster R-CNN [9], and the SSD method [10] by framing object detection as a regression problem that takes into account spatially separated bounding boxes and their associated class probabilities for faster and better real-time object detection and classification with localization through bounding boxes around it. Hence, we exploit the YOLO-based method for guide dog robots.

Commercial quadrupedal robots are becoming more affordable as the industry expands, and they can be used as a base platform for robotic guide dog technology. In this work, we propose a robotic guide dog sensing system combined with a deep learning algorithm for real-time indoor object perception. The main contributions are four-fold. They include (a) the development of a sensing module and its integration with a quadrupedal robot to obtain real-time visual data; (b) the development of the YOLO-GUIDE framework for real-time object detection and multi-label classification with localization; (c) the demonstration of the proposed method's high performance through qualitative and quantitative evaluations; and (d) examined the YOLO-GUIDE's performance against other relevant models and shown its superiority.

II. Methodology

A. Development of a Robotic Guide Dog Sensing System

In this study, we used a commercially available quadrupedal robot (SPOT, Boston Dynamics) as a base platform for developing a robotic guide dog. Since the base robot only had stereo cameras, which provided monochromatic visual data input, our first algorithmic developments revealed poor performance for object perception. Therefore, developing a custom sensor module is necessary to develop guide dog robots. In this circumstance, a sensor module as shown in Fig. 1(a) was developed. The sensor module comprises a Raspberry Pi 4 Model B-based single-board computer (SBC) with 16 GB of RAM and a 5-megapixel color camera (Raspberry Pi Camera). To mount the SBC and hold the camera, a polylactide material casing was 3D printed. The casing was then screwed into the robotic guide dog platform as shown in Fig. 1(b), so that it gets a clear view of the surroundings in front of the robot. The robot provided power by utilizing a specially designed adapter for the robot's DB25 connection since the SBC required 5V and at 3A. The sensor module was powered, connected to a local WiFi network, and set up to communicate with the camera to transmit real-time video data for processing. Any other device on the same network can get the video through an address since it is encoded using the H.264 standard. Following the start of the video transmission, another computer connected to the same local network can receive the data and utilize it as the input media source for the YOLO-GUIDE framework.

B. YOLO-GUIDE Framework Development

Figure 2 presents an illustration of the YOLO-GUIDE framework, where the indoor environment scene is captured by the robot and processed by the YOLOv7 [11] architecture for object perception, such as detecting the objects, multi-label classification with localization of objects in a given data source. Image frames in a YOLO model are enhanced by a backbone. The extended efficient layer aggregation network (E-ELAN) architecture that makes up the YOLOv7 backbone functions as a feature extractor from the input data source. The Neck essentially acts as a feature aggregator by gathering feature maps from several backbone phases. The head, which is referred to as the object detector of the network, receives the properties after they have been integrated and blended in the neck of the network, which then predicts the locations and categories of objects around which bounding boxes ought to be created. The object localization process involves putting a box around the classified objects. Ideal parameters for training YOLO-GUIDE were 300 epochs and a batch size of 32.

III. Experimentation and Results

A. Data Set and Computation Tools

The publicly available COCO data set [12] was used to access the YOLO-GUIDE framework. In this work, ten distinct classes of objects often present in an indoor environment were chosen for this robotic guide dog application. (i) bed, (ii) cell phone, (iii) chair, (iv) couch, (v) dining table, (vi) laptop, (vii) potted plant, (viii) television (TV), (ix) person, and (x) dog are the ten object classes. The YOLO-GUIDE was trained using 75,000 images and tested with 5,000. We implemented YOLO-GUIDE in the UTS Cluster that consists of 30 nodes running on RHEL 8.6 with an Intel Xenon Gold 6238R 2.2GHZ 28 cores 38.5MB L3 Cache, 180GB RAM, 3.5TB of scratch disk space, GPU - NVIDIA Quadro RTX 6000 Passive.

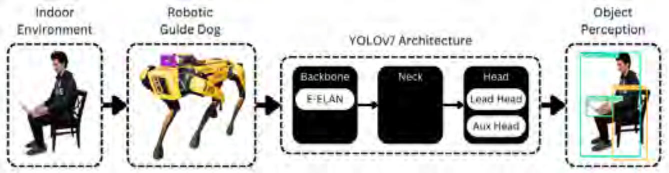


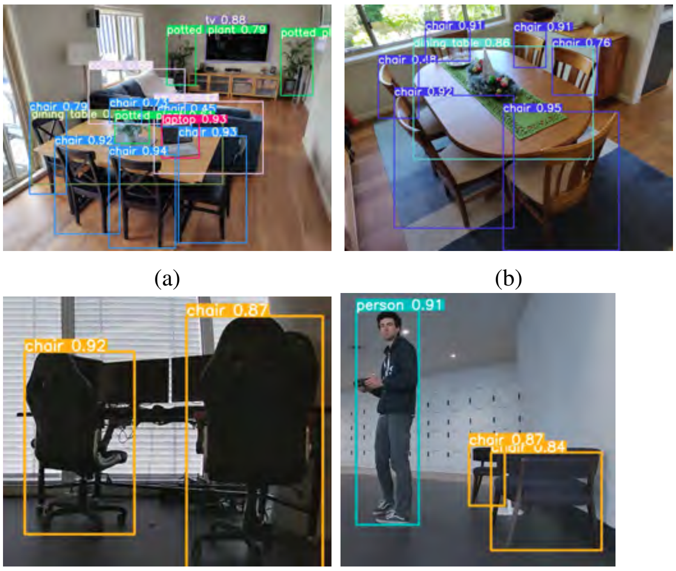
Fig. 2: Illustration of YOLO-GUIDE framework.

B. Qualitative Evaluation of the YOLO-GUIDE Framework

A qualitative assessment of the YOLO-GUIDE framework was conducted to determine whether it delivers the intended results. Real-time videos were captured in indoor settings, such as homes and offices. The guide dog robot feeds real-time video into the YOLO-GUIDE framework, which outputs rapid feedback by displaying detected and classified objects with localization through a bounding box around each object detected in the frame and a confidence score next to the class name as in Fig. 3. The major visible indicator for evaluating the framework's performance in real-time is the confidence score; as the value gets closer to 1.00, the framework is more confident that its prediction is accurate. It was encouraging to observe that, with the exception of a few obscure, and challenging-to-see things, the vast majority of the objects in the environment were detected and correctly classified with multi-labels and localization as in Fig. 3, even though certain confidence values are still low. It was observed in a few instances that the reflection of the person on the glass door could be mistaken for a person's class as in Fig. 4(a), and the camera held on hand was detected as a cell phone class as in Fig. 4(b). The next project phase will solve this problem by having other complementary sensing modalities to counteract such inaccurate detection. Overall, it can be stated that there were many instances of confidence scores that were between 0.8 and 0.9, which is an indicator of high performance as seen in Fig. 3. Hence, the qualitative evaluation was satisfactory.

C. Quantitative Evaluation of the YOLO-GUIDE Framework

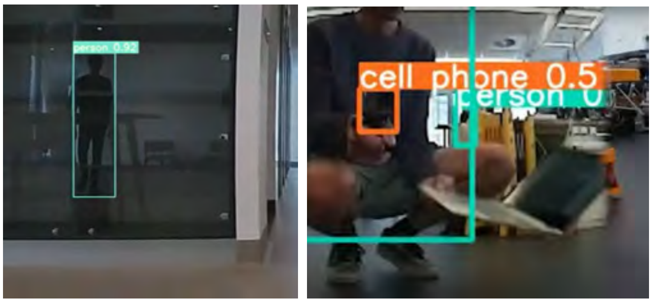
The F1-score, which is the harmonic mean of Precision and Recall, is the most often used statistical measure for evaluating object detection and classification ability. When applied to imbalanced data, such as the dataset used for this study, the F1-score offers a more realistic depiction of a classification framework's performance [13], [14]. The F1-scores have a scale of 0 to 1, with 1.0 denoting perfect Recall and Precision. Precision is the fraction of relevant instances found among the retrieved instances, whereas Recall is the fraction of relevant instances found. After a YOLO-GUIDE



(a) (b)

(c) (d)

Fig. 3: Qualitative evaluation of multi-label classification with the object localization and confidence score. (a) Living room, (b) Dining area, (c) Office desk space, (d) Office lounge area.



(a) (b)

Fig. 4: Incorrect classifications. (a) A person's reflection on a glass door is incorrectly classified as person class, and (b) A person holding a camera was classified as cell phone class.

framework has been trained, the F1-scores are calculated using the resulting Precision and Recall values from each training epoch. In all, the YOLO-GUIDE framework earned 0.76 Precision, 0.67 Recall, and an F1-score of 0.71, which is high performance for the 10 objects shown in Table I.

D. Performance Comparison

Numerous studies demonstrate that YOLO-based object detection and multi-label object classification with localization outperform other cutting-edge models in terms of detection speed and accuracy [15], [16]. Therefore, in this work, we compare the performance of several kinds of recent YOLO-based algorithms such as YOLOv5, YOLOv6, and YOLOv8 with the proposed YOLO-GUIDE. The 2020 release of YOLOv5 by Ultralytics [17] makes use of an EfficientDet [18], which enables YOLOv5 to outperform earlier YOLO models in terms of accuracy. [19] proposed YOLOv6 in 2022, which used EfficientNet-L2. It has fewer parameters and a more efficient computational model than the YOLOv5. The YOLO-GUIDE framework is based on YOLOv7, which employs nine anchor boxes and can identify a larger variety of object shapes and sizes than earlier YOLO versions. Faster processing speed makes YOLOv7 appropriate for guide dog robots. YOLOv8 was released by Ultralytics [17] in December 2022. The YOLOv5, YOLOv6, and YOLOv8 models were trained and tested similarly to the YOLO-GUIDE framework in ideal environment settings. The performance metrics are reported in Table 2, where it can be shown that the proposed YOLO-GUIDE framework performs better than others in terms of the F1-score, demonstrating its efficacy. The Precision of the YOLO-GUIDE is 0.77, 0.06 more than that of the YOLOv5, 0.02 higher than that of the YOLOv6, and 0.03 greater than that of the YOLOv8. YOLO-GUIDE's Recall is 0.66, which is comparable to YOLOv6 but 0.09 more than YOLOv5 and 0.06 better than YOLOv8. From this comparison, it can be stated that the proposed YOLO-GUIDE demonstrated its superiority over the other examined models for indoor object perception.

TABLE I: YOLO-GUIDE Computed Performance Metrics.

Object Classes	Precision	Recall	F1-score
Laptop	0.88	0.85	0.87
Dog	0.85	0.81	0.82
Person	0.84	0.78	0.80
TV	0.81	0.75	0.78
Bed	0.80	0.75	0.77
Couch	0.69	0.65	0.68
Dining Table	0.73	0.54	0.63
Cell Phone	0.71	0.55	0.63
Potted Plant	0.64	0.50	0.56
Chair	0.68	0.48	0.56
Overall	**0.76**	**0.67**	**0.71**

TABLE II: Performance Comparison.

Methods	Precision	Recall	F1-score
YOLOv5	0.71	0.57	0.63
YOLOv6	0.75	0.66	0.70
YOLOv8	0.74	0.60	0.66
YOLO-GUIDE	**0.77**	**0.66**	**0.71**

IV. Conclusion And Future Work

This work proposed a robotic guide dog sensing system for real-time indoor object perception for VIPs. A sensor module was created and linked to the YOLO-GUIDE, allowing it to detect and classify objects with localization while the robot navigates in its indoor settings. Video of real-time implementation is in [20]. Quantitative assessments show that the proposed framework works well in most situations, and quantitative evaluation results show that the proposed method performs effectively, with a Precision of 0.76, Recall of 0.67, and F1-score of 0.71 for ten object classes. The YOLO-GUIDE framework outperformed other models, indicating its superiority. In the future, YOLO-GUIDE will add more data classes and estimate object distances.

References

[1] F. E.-Z. El-Taher, L. Miralles-Pechuan, J. Courtney, K. Millar, C. Smith, and S. Mckeever, “A survey on outdoor navigation applications for people with visual impairments,” *IEEE Access*, vol. 11, pp. 14 647–14 666, 2023. [Online]. Available: https://doi.org/10.1109/access.2023.3244073

[2] S. Bhatlawande, S. Shilaskar, A. Kumari, M. Ambekar, M. Agrawal, A. Raj, and S. Amilkanthwar, “AI based handheld electronic travel aid for visually impaired people,” in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*. IEEE, Apr. 2022, pp. 1–5. [Online]. Available: https://doi.org/10.1109/i2ct54291.2022.9823962

[3] A. Sen, K. Sen, and J. Das, “Ultrasonic blind stick for completely blind people to avoid any kind of obstacles,” in *2018 IEEE SENSORS*. IEEE, Oct. 2018, pp. 1–4. [Online]. Available: https://doi.org/10.1109/icsens.2018.8589680

[4] E. Cardillo, V. D. Mattia, G. Manfredi, P. Russo, A. D. Leo, A. Caddemi, and G. Cerri, “An electromagnetic sensor prototype to assist visually impaired and blind people in autonomous walking,” *IEEE Sensors Journal*, vol. 18, no. 6, pp. 2568–2576, Mar. 2018. [Online]. Available: https://doi.org/10.1109/jsen.2018.2795046

[5] S. Bhatlawande, M. Mahadevappa, J. Mukherjee, M. Biswas, D. Das, and S. Gupta, “Design, development, and clinical evaluation of the electronic mobility cane for vision rehabilitation,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 6, pp. 1148–1159, Nov. 2014. [Online]. Available: https://doi.org/10.1109/tnsre.2014.2324974

[6] K. Thiyagarajan, S. Kodagoda, M. Luu, T. Duggan-Harper, D. Ritchie, K. Prentice, and J. Martin, “Intelligent guide robots for people who are blind or have low vision: A review,” *Vision Rehabilitation International*, vol. 13, pp. 1–15, 2022. [Online]. Available: https://sciendo.com/article/10.2478/vri-2022-0003

[7] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2016, pp. 779–788. [Online]. Available: https://doi.org/10.1109/cvpr.2016.91

[8] S. S. Sumit, J. Watada, A. Roy, and D. Rambli, “In object detection deep learning methods, YOLO shows supremum to mask r-CNN,” *Journal of Physics: Conference Series*, vol. 1529, no. 4, p. 042086, Apr. 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1529/4/042086

[9] T. Mahendrakar, A. Ekblad, N. Fischer, R. White, M. Wilde, B. Kish, and I. Silver, “Performance study of YOLOv5 and faster r-CNN for autonomous navigation around non-cooperative targets,” in *2022 IEEE Aerospace Conference (AERO)*. IEEE, Mar. 2022, pp. 1–2. [Online]. Available: https://doi.org/10.1109/aero53065.2022.9843537

[10] M. Li, Z. Zhang, L. Lei, X. Wang, and X. Guo, “Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster r-CNN, YOLO v3 and SSD,” *Sensors*, vol. 20, no. 17, p. 4938, Aug. 2020. [Online]. Available: https://doi.org/10.3390/s20174938

[11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YoLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2023, pp. 7464–7475. [Online]. Available: hhttps://openaccess.thecvf.com/content/CVPR2023/papers/Wang_YOLOv7_Trainable_Bag-of-Freebies_Sets_New_State-of-the-Art_for_Real-Time_Object_Detectors_CVPR_2023_paper.pdf

[12] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, and P. Dollár, “Microsoft coco: Common objects in context,” 2014. [Online]. Available: https://arxiv.org/abs/1405.0312

[13] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, “Optimal thresholding of classifiers to maximize f1 measure,” in *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, 2014, pp. 225–239. [Online]. Available: https://doi.org/10.1007/978-3-662-44851-9_15

[14] X. Wang, K. Thiyagarajan, S. Kodagoda, and M. Zhang, “PIPE-CovNet: Automatic in-pipe wastewater infrastructure surface abnormality detection using convolutional neural network,” *IEEE Sensors Letters*, vol. 7, no. 4, pp. 1–4, Apr. 2023. [Online]. Available: https://doi.org/10.1109/lsens.2023.3258543

[15] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, “Comparative analysis of deep learning image detection algorithms,” *Journal of Big Data*, vol. 8, no. 1, May 2021. [Online]. Available: https://doi.org/10.1186/s40537-021-00434-w

[16] R. Cheng, “A survey: Comparison between convolutional neural network and YOLO in image identification,” *Journal of Physics: Conference Series*, vol. 1453, no. 1, p. 012139, Jan. 2020. [Online]. Available: https://doi.org/10.1088/1742-6596/1453/1/012139

[17] G. Jocher, A. Chaurasia, and J. Qiu, “YOLO by Ultralytics,” Jan. 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[18] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2020, pp. 10 778–10 787. [Online]. Available: https://doi.org/10.1109/cvpr42600.2020.01079

[19] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, “Yolov6: A single-stage object detection framework for industrial applications,” pp. 1–17, 2022. [Online]. Available: https://arxiv.org/abs/2209.02976

[20] 2023. [Online]. Available: https://www.youtube.com/watch?v=PCxe-lljm8U