UTM UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Object Detection Algorithms for Autonomous Navigation Wheelchairs in Hospital Environment

Mohamed Mokhtar Ouardi[1] & Dayang N. A. Jawawi[2]
Faculty of Computing,
Universiti Teknologi Malaysia
81310, UTM Johor Bahru, Johor, Malaysia
Email: mokhtar.ouardi@yahoo.fr[1]; dayang@utm.my[2]

Submitted: 17/1/2024. Revised edition: 28/3/2024. Accepted: 12/5/2024. Published online: 31/5/2024
DOI: https://doi.org/ 10.11113/ijic.v14n1.459

***Abstract*—Under the overwhelming situation medical facilities struggle to provide sufficient personnel to assist the continuous traffic. Therefore, considering alternative solutions to help nurses transport patients is relevant. An autonomous navigation wheelchair driven by object detection can become a substitute for transporting patients between facilities. The objective of this research is to study the potential of object detection algorithms in facilitating the autonomous navigation of patients within hospitals. For an object detection model to be a standalone technology for driving the autonomous wheelchair it has to satisfy the standards of the domain and display adequate readiness for deployment. Precisely, conduct viability experimentation based on Efficiency, Safety, and Reliability on a You Only Learn One Representation (YoloR) model trained for the research purpose. The paper finds YoloR as an appropriate model with potential for deployment. Both in terms of Reliability and Safety, the algorithm fits the designated criteria. However, the Autonomous Wheelchair system is a real-time system with strict execution time requirements that YoloR on its own does not reach. The latter is further confirmed by the proposed testing method.**

***Keywords*—Artificial Intelligence, Object detection, You Only Learn One Representation, Robotic wheelchair**

I. INTRODUCTION

Object detection has been a fundamental tool in automated computer vision systems. Particularly with the advancement of machine learning algorithms, the detection tools have proved their capability in speeding up the detection models that structure the system. Object detection is a challenging task that involves predicting both where the objects are in the image and what type of objects were detected which requires identifying behaviors, and characteristics. These attributes when initialized by humans are hard to grasp from the context of a computer. This is where Artificial Neural Network trained models to come to fruition. Indoor autonomous navigation is a necessity in the circumstances of an overloaded workspace. Safe and reliable transport systems that do not require human intervention are important in a hospital environment. The latter suffers from restricted proficient human resources which call for high efficiency and availability of the core medical service. Therefore, a wheelchair driven by object detection can become a substitute for transporting patients between facilities.

A robotic wheelchair is a motorized chair with embedded hardware and software that is designed to facilitate transport for patients with mobility limitations. Additionally, when exploring autonomous navigation wheelchair literature, the major driving technologies are depth-data sensors coupled with object detection models. Systems achieve accurate positioning by building a large-scale 3D map which is both computationally expensive and hardly scalable due to the hardware technology required. While Object detection algorithms based on deep learning are a commodity in current technology it is necessary to establish a testing process to ensure the viability of the model. To demonstrate relevant testing experimentation a requirements-based testing strategy must be established. For this reason, the improved string distance through considering requirements priority weight [1] served as the template for the proposed focus on testing the algorithm while arbitrarily setting weights to the model base metrics to create a scoring system suitable for model contextual evaluation. Viability is the ability of the object detection model to achieve the computer vision task while satisfying the requirement established by the context. Factoring that Autonomous Navigation Wheelchair is a real-time system subject to strict criteria for the safety of the passenger. Under the context of medical facilities, the requirements are enhanced to suit ISO 9001 which dictates quality and risk management. Therefore, under these conditions, the research proposes three contextual metrics Efficiency, Reliability, and Safety which ensure the compatibility of the system with hospitals and its readiness for deployment. The paper aims to discuss the viability of object detection for indoor navigation in the context of hospital facilities by proposing testing experimentation that factors in the context. Ultimately the research tries to answer how viable convolutional neural networks object detection models are for indoor autonomous navigation and whether YoloR is ready for deployment in a hospital environment.

II. RELATED WORKS

Looking into the literature, a pattern emerges. Rarely is object detection used independently for navigation. The majority of papers explore algorithms and technologies such as Rapid Random Tree (RRT) path planning or Lidar ROS coupling. These methods are dependent on localization. Location information is essential and they are essentially retrieved through sensors. The issue lies in the uncertainty in both the driving and the sensing of the robot which decreases the reliability of the system. Additionally, accessibility to end sensors can be restrictive for the scalability and availability of the solution.

A. Object Detection and Tracking for Autonomous Navigation in Dynamic Environments

The study addresses the issues of vision-based navigation. The author categorizes the challenge as a geometric accuracy restriction. They underline a comparison between LIDAR sensors and Digital cameras where they highlight the restricted amount of data generated by the latter. For instance, environment modeling for robot navigation requires dynamic space localization which is complex without depth data. With this intention, the research proposes an object detection algorithm supported by trajectory estimation over time [2]. The experimentation supports the reliability of object detection systems for navigation purposes in busy and dynamic scenarios.

B. Autonomous Navigation Research for Mobile Robot

The study tackles the bionic strategy to operant conditioning mechanism. In particular, the project defines a method to solve navigation issues in an unknown environment which is based on reinforcement and operant conditioning learning. The strategy establishes a simulation of sensor data feedback entropy which influences the tendency degree [3]. The Navigation trajectory based on a bionic autonomous learning strategy has quick learning velocity and accurate navigation ability. However, the strategy is generally designed to evade static obstacles and reach a known goal point.

C. Object Detection Learning Techniques for Autonomous Vehicle Applications

The work investigates learning models for video-based object detection applied to self-driving vehicles. The author studies the potential application of three object detection techniques in achieving autonomous navigation: Object detection with the Support Vector Machine (SVM) model, with YOLO, and with SSD. The author proposed a comparison method where they subdivide video frames into classes to compare the three algorithms over processing speed and detection accuracy [4]. Ultimately, the author concluded that YOLO should be employed for extremely real-time processing due to its rapidity in treating the frame while SSD can be employed for its high accuracy in detecting small objects.

D. Object detection and tracking using faster r-CNN

The work [5] studies the influence of hyperparameters on the output model performance. The impacts of batch size, number of iterations, and learning rate are investigated. The author discusses the trade-off speed to accuracy when elaborating on their algorithm choices. Their findings show an increase of accuracy of 5% through hyperparameters tuning. While the processing speed was non-correlated with the operation.

TABLE I. COMPARISON BETWEEN EXISTING PAPER AND PROPOSAL

Research Work	**Domain**	**Quality Criteria**	**Algorithm**
(Ess, Andreas & Leibe, 2010) [6]	Computer Vision	Accuracy	EKF-trackers
(J. Cai, R. Yu, 2012) [7]	ROS / automated machine learning	Reliability	Bionic Autonomous Learning
(M. Masmoudi, H. Ghazzai, 2019) [8]	Computer Vision	Efficiency	SVM / YOLO / SSD
(Chakradhara Panda, 2019) [9]	Computer Vision	Efficiency	Faster R-CNN / SSD / YOLO
Proposed Research	Computer Vision	Safety, Reliability, Efficiency	EfficentDet / YoloR

After studying the previous papers, the data from each paper are collected and tabulated as shown in Table I. The keys from each paper that has been recorded are project domain, quality criteria, and algorithm. A comparison is designated to contribute to this research's objective. This work focuses on implementing machine learning-driven object detection for navigation. On this matter, the domain of study has been established as Computer vision. When elaborating on quality criteria the context of the research plays a major role.

III. METHODOLOGY

To apply the most suitable object detection algorithm on the Autonomous Navigation Wheelchair and perform simulation testing in a Hospital Environment. YoloR fits the requirements set for real-time systems. The latter displays a relatively stable accuracy with potential for high-speed processing. Accordingly, personalized contextual metrics have to be defined. Efficiency is derived from the Precision and speed of the model. Reliability from Precision, Recall, mAP and confidence. Safety from mAP and Confidence. These scores are dedicated to confirming the validity of the model in the context of the research.

Fig. 1 displays the strategy established for the experimentation. The flow starts with data preparation which is collected between 2 case studies UTM Health Center and Hospital Sultanah Aminah which provide great variance within the dataset. While the Hospital dataset offers an accurate representation of the deployment domain, the UTM Health Center Dataset introduces simulated scenarios especially manufactured to test the model under unexpected situations. The second step consists of running the YoloR inference and collecting the base metrics. These metrics are then used to extract the contextual scores and evaluate the performance of the model.

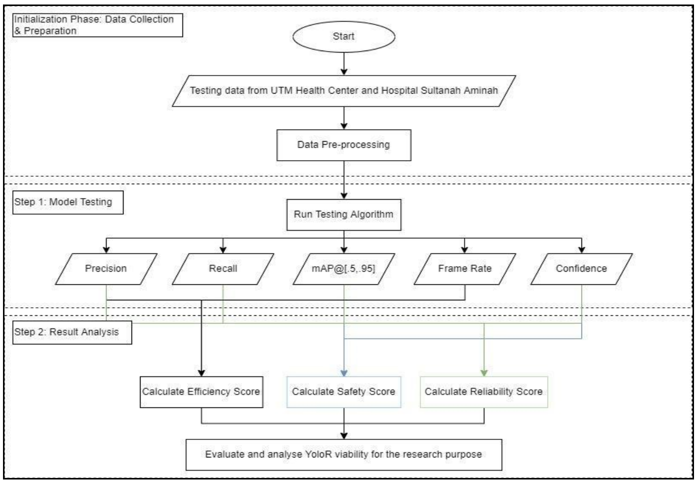

Fig. 1. YoloR Testing Overview

A. Experiment Design

The testing is based on three contextual metrics Efficiency, Reliability, and Safety. The objective of these metrics is to ensure compatibility with ISO 9001 for hospitals. Considering the absence of suitable testing methods a scoring strategy is proposed. The scoring is extrapolated from the state-of-the-art testing processes [10]. The testing metrics are weighted to generate a score depending on the key features. A threshold is arbitrarily set up based on a literature review coupled with ISO criteria. Table II encapsulates the determined thresholds and the proposed equation used for the score calculation.

TABLE II. CONTEXTUAL METRIC WEIGHTS

	Threshold	**Scoring Equation**
Precision	85%	$\Sigma\{(\text{Result / Threshold}) * \text{Weight}\}$
Recall	90%	
mAP	95%	
Confidence	50%	
Speed (fps)	30	

i. The recall is the ratio of the number of true positives to the total number of relevant objects.
ii. Precision is the ratio of the number of true positives to the total number of positive predictions.
iii. The mAP is a ratio of the ground-truth bounding box to the detected box.
iv. The confidence score reflects how likely the box contains an object and how accurate the bounding box is.
v. The speed is the frame rate per second indicating how fast each frame generates the detection results.

TABLE III. CONTEXTUAL METRIC WEIGHTS

	Precision	**Recall**	**mAP**	**Confidence**	**speed**
Precision	0	0	1	0.5	0
Recall	0.8	1	1	0.5	0
mAP	0.5	0	0	0	1

Table III indexes the proposed weights assigned to each base metric regarding the contextual metrics. The weights are arbitrarily selected and extrapolated from the literature review of [10] which is a Survey on Performance Metrics for Object-Detection Algorithms coupled with ISO 9001. Safety is a representation of the risk involved in the implementation of the model for autonomous navigation. Therefore, the latter is extrapolated from the mean average precision which is the major indicator of the model detection quality and the confidence which is an indication of the labeling validity. Evidently, due to the limited risk involved with miss labeling operations which are mainly confusing between a static and dynamic object the weight of Confidence is only 50%. Reliability is a representation of the fundamental accuracy of the model. This contextual metric is defining the ability of the model to interact with the environment of a hospital while taking into account false detections. Under this definition, the Reliability score is mainly based on Recall and mAP which both underline the accuracy and the tendencies for irrelevant detections. Coupled with the latter are Confidence and Precision both as an indication of the odds for false negatives and miss labeling. The weights are justified by judging the prospect of false calls compared to escapees. In this context, an escapee can lead to endangering the user by crashing into an obstacle while a false call will only stop the system from moving. Efficiency is the metric for real-time system implementation. It is essential to understand how the algorithm behaves in a dynamic environment. Identifying the probability of missing detection frames during the process and their implication in the system functionality. Therefore, the metric is calculated from the processing speed coupled with the precision of the model.

B. YoloR Conformity Testing

YoloR architecture offers a single neural network that performs multiple parallel convolutional tasks. This concept targets acute feature alignment, prediction refinement, and multi-task learning [11]. Performing object detection, multi-label image classification, and feature embedding as aligned tasks push the model to extract complex feature extraction [12]. The extracted features are favorable for better generalization and refined model attributes. Ultimately, the prediction refinement provided by YoloR enhances the model by performing correction and fine-tuning using the loss function. YoloR model when compared to YoloV4 is 2% more accurate with approximately 300% gain in processing speed [13]. Different approaches have been employed to solve the growing need for accurate object detection models. Once the model is trained it becomes essential to identify the suitable method compatible with the implementation's context. The necessity for an objective testing process and environment is essential to estimate the performance in the field. Therefore having testing set up as close as the real-world implementation is fundamental [1]. Essentially, for the research purpose, a demo robot with a similar navigation system recording through the medical facilities destined for the Autonomous Navigation Wheelchairs could be the qualification criteria. Hence, bypassing the standard results on the testing dataset the model could be considered ready for implementation. This operation can also highlight limits that can be addressed before deployment.

C. Testing Environment

The robot represented in Fig. 2 is equipped with a UTM Robocon personalized navigation board along with the Single Board Computer Jetson Nano for object detection and dataset recording. Additionally, a smartphone is a secondary camera for data collection operations. The testing setup was dispatched to Hospital Sultanah Aminah and UTM Health Center for two data collection operations during active working hours.

Fig. 2. Testing Robot

IV. TESTING AND RESULT

The implementation of the YoloR model in the Hospital dataset produces relatively good output. This behavior displays a strong generalization inherited from the ancestor weights. This feature carries into the ability to handle occlusion and noise allowing a stable detection and tracking of the dynamic objects. Fig. 3 shows that the model is capable of detecting small/far objects which can be an asset to predicting and reacting to an obstacle before critical range.

Referring to Fig. 4 the testing algorithm displays an mAP@[.5,.95] which is the mean average precision at different Intersections over Union percentages, of 88.6%. A precision of 89.3% and a recall of 98.6%. The time elapsed by inference is equal to 41.3 ms which gives 24 frames per second.

The YoloR model testing on the real hospital dataset serves as an indicator of the model quality and potential viability in a concrete scenario. When factoring both the results displayed in Figs. 2 and 3, the model is determined to produce adequate multi-image labeling and object detection at a relatively sustainable processing speed for embedded systems such as an autonomous wheelchair. Although, the generalization offered by retraining the base model generates unnecessary detections. This behavior only produces false calls which are minor inconveniences that only lead to the system being extra safe during navigation. Table IV tabulates the metrics collected from the testing algorithm.

Fig. 3. Test Detection Output

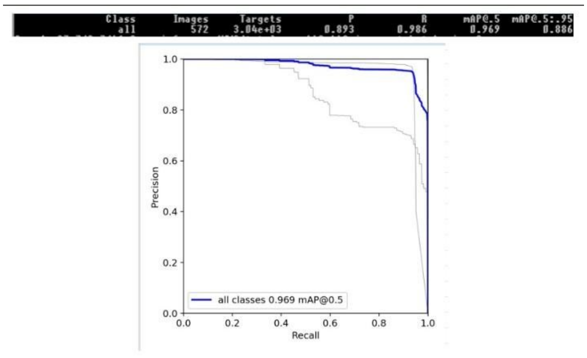

Fig. 4. Test Detection Results

TABLE IV. TABULATED BASE METRIC RESULTS

	Threshold	Results
Precision	85%	89,3%
Recall	90%	98,6%
mAP	95%	88,6%
Confidence	50%	63%
Speed (fps)	30	24

V. ANALYSIS

From a metric standpoint, YoloR is producing convincing results. While 0.886 mAP@[.5,.95] on the contextual dataset is still weak for real-world deployment, the potential for improvement after model optimization and further fine-tuning with contextual data could push the accuracy further.

However, the speed of the model is concerning. When factoring that the SBC was 100% dedicated to object detection and not driving functions, the 24 fps becomes a drastic risk. As the autonomous wheelchair functionalities are highly dependent on its real-time system capabilities any frame loss or missed detection might lead to critical failures which are not acceptable in the context of the research.

When converting the base metric results to the proposed testing score as demonstrated in Table V, YoloR barely satisfies the Safety and Reliability. This translates to the potential prospects of the model for a 100% object detection-driven autonomous navigation system. However, it denotes some reasonable limitations. Object detection models including YoloR are vulnerable to "unknown" objects that have not been introduced. Without explicit supervision, YoloR is unable to adapt which could drive the risk factor to unreasonable levels for deployment. Furthermore, YoloR fails the Efficiency criteria which is a concerning result for the YoloR viability for the research context. The possibility of missing frames for an autonomous navigation system could lead to delayed reactions and inconsistent navigation.

TABLE V. TABULATED PROPOSED SCORE RESULTS

	Target*	**Score**	**Δ***
Safety	1.5	1.562	+0.062
Reliability	3.3	3.49	+0.19
Efficiency	1.5	1.32	-0.18

Target* = Σ{(Threshold / Threshold) * Weight}
Δ* = The difference between the target and score

Yolo is a state-of-the-art model designed to push the limit of object detection. However, from a real-time system perspective, the model requires enhancement to be industry ready. The outcome denotes the potential of YoloR for autonomous navigation. However, obvious limitations such as processing speed shortage and unknown object detection make technology still weak to become a stand-alone in the research application. Therefore, YoloR can be a base for integrating and establishing models. The model still requires improvement to be eligible for the autonomous navigation wheelchair in a hospital environment. From the domain perspective, the model weakness brings drastic risks that are not sustainable for the system's functionality. Despite the relatively high accuracy output the potential for detection failures due to computation limitations is too unpredictable to be trusted with the safety of patient transport. For such technologies to gain the trust of a strict field such as a hospital the system would need to be flawless which is currently not reachable without support technologies such as Lidar.

Testing on contextual metrics such as Efficiency, Safety, and Reliability encapsulates the requirements that object detection models need to satisfy to be eligible for real-world implementation. This strategy factors in the standards and criteria involved in medical facilities. While this method is dedicated to the medical field in this paper it can serve as a template to establish contextual metrics necessary to evaluate a model for deployment in the industry.

VI. CONCLUSION

Ultimately, the model displays the ability for obstacle detection in medical facilities. The model manages the crowded environment and is capable of differentiating between static and dynamic objects. It is also able to recognize humans with great precision. However, the restrained YoloR is vulnerable to "unknown" objects that represent potential missed obstacles. Furthermore, the model is not fully ready for real-time deployment. The latter still requires enhancement and optimization to reach the required inference time necessary for the deployment. YoloR shows potential for navigation wheelchair in the hospital environment. Nevertheless, in the current state, the model still requires the assistance of supporting technologies such as Lidar to be valid for a fully autonomous system.

However, future works proposing Open World object detection algorithms to incrementally learn unknown categories while holding to previously learned classes could be a potential improvement. This concept could drastically boost the viability of object detection models in standalone operations such as for research purposes. At the same time to satisfy the healthcare real-time requirements, optimization to increase the inference time can be considered. Furthermore, additional Pruning which is a method of removing redundant connections present in the architecture, and Weight and Activation Precision Calibration could improve the model speed.

ACKNOWLEDGMENT

We would like to thank the authority of Universiti Teknologi Malaysia (UTM) for providing us access to their facilities for research purposes. Likewise, We would like to extend our appreciation to Hospital Sultanah Aminah and UTM Robocon for their collaboration and the opportunities offered.

REFERENCES

[1] Dayang, N. A. Jawawi, Rooster Tumeng, & M. Irsyad Kamil R. (2021). Requirements-based testing for robotic wheelchair tracking software system. *Malaysian Journal of Medicine and Health Sciences, 17*(SUPP 13), 47-53.

[2] Ess, Andreas & Leibe, Bastian & Van Gool, Luc. (2010). Object detection and tracking for autonomous navigation in dynamic environments. *I. J. Robotic Res.*

[3] J. Cai, R. Yu and L. Cheng. (2012). Autonomous navigation research for mobile robot. *Proceedings of the 10th World Congress on Intelligent Control and Automation,* 331-335.

[4] M. Masmoudi, H. Ghazzai, M. Frikha and Y. Massoud. (2019). Object detection learning techniques for autonomous vehicle applications. *IEEE International Conference on Vehicular Electronics and Safety (ICVES).*

[5] Chakradhara Panda. (2019) Object detection and tracking using faster r-cnn. *International Journal of Recent Technology and Engineering*, *8*(3), 4894-4900.

[6] Ess, A., Schindler, K., Leibe, B., Van Gool, L. (2010). Object detection and tracking for autonomous navigation in dynamic environments. *The International Journal of Robotics Research, 29*(14), 1707-1725. Doi:10.1177/0278364910365417.

[7] J. Cai, R. Yu and L. Cheng. (2012). Autonomous navigation research for mobile robot. *Proceedings of the 10th World Congress on Intelligent Control and Automation, Beijing, China,* 331-335. Doi: 10.1109/WCICA.2012.6357893.

[8] M. Masmoudi, H. Ghazzai, M. Frikha and Y. Massoud. (2019). Object detection learning techniques for autonomous vehicle applications. *2019 IEEE International Conference on Vehicular Electronics and Safety (ICVES), Cairo, Egypt,* 1-5. Doi: 10.1109/ICVES.2019.8906437.

[9] Panda, C. (2019). Object detection and tracking using faster R-CNN. *International Journal of Recent Technology and Engineering (IJRTE).*

[10] Padilla, R., Netto, S. L., & da Silva, E. A. (2020). A survey on performance metrics for object-detection algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP),* 237-242.

[11] Wang, C. Y., Yeh, I. H., & Liao, H. Y. M. (2021). You only learn one representation: Unified network for multiple tasks. arXiv preprint arXiv:2105.04206.

[12] Tran, V. T., To, T. S., Nguyen, T. N., & Tran, T. D. (2022). Safety helmet detection at construction sites using YOLOv5 and YOLOR. *International Conference on Intelligence of Things.* Cham: Springer International Publishing. 339-347.

[13] Wang, C.-Y., Mark Liao, H.-Y., & Yeh, I.-H. (2021). You only learn one representation: unified network for multiple tasks. arXiv preprint arXiv:2105.04206.