Deep Transfer Learning with Convolutional Neural Networks for Object Detection and Recognition in Autonomous Mobile Robots

Onah Bartholomew Chukwuebuka
Department of Computer Science, Faculty of Technology, Institute of Management and Technology, PMB 01079, Independence Layout, Enugu, Enugu State, Nigeria

* Corresponding Author: **Onah Bartholomew Chukwuebuka**

Article Info

ISSN (Online): 3107-6580
Impact Factor (RSIF): 8.23
Volume: 02
Issue: 05
Received: 06-07-2026
Accepted: 04-08-2026
Published: 02-09-2026
Page No: 01-09

Abstract
Autonomous mobile robots (AMRs) are increasingly deployed across industrial automation, logistics, healthcare, surveillance and hazardous environments. Their effectiveness depends on the ability to detect and recognise obstacles reliably in unstructured environments. Conventional convolutional neural networks (CNNs) trained from scratch require large annotated datasets and long training times, which limits deployment on resource-constrained mobile platforms. This work studies the modelling of a four-wheel holonomic autonomous mobile robot and applies a deep transfer learning (DTL) algorithm based on a fine-tuned AlexNet CNN for obstacle detection and recognition. The Dynamic Systems Development Model (DSDM) was adopted, and the algorithm was implemented and validated in Simulink/MATLAB Robotics Operating System (ROS). Experimental results across ten test runs yielded a mean recognition accuracy of 98.7% (0.96 s mean detection delay and 97.7% manoeuvre success rate), representing a 1.89% improvement over the baseline CNN benchmark and outperforming several state-of-the-art comparators.

DOI: https://doi.org/10.54660/.IJECA.2026.2.5.01-09

Keywords: Autonomous mobile robot, Deep transfer learning, Convolutional neural network, AlexNet, Obstacle detection, Simulink/ROS

1. Introduction

Over the last six decades, robotic technology has revolutionised the industrial sector, providing a redundancy layer that facilitates technical processes and maintains production standards. Robots can perform complex and dangerous tasks repeatedly, execute multiple functions tirelessly, and operate accurately at high speed and low cost. Consequently, robotics has become one of the fastest expanding fields of scientific research, with applications ranging from surveillance, planetary exploration, patrolling, emergency rescue, reconnaissance, petrochemical operations, industrial automation, construction, entertainment, museum guidance, personal services, extreme-environment intervention, transportation and medical care (Rubio *et al*., 2019) [11].

Robots are generally classified as standalone, autonomous, or mobile (Surmann *et al*., 2020) [13]. The first-generation standalone robot (for example, the traditional robotic arm or manipulator) is designed with several degrees of freedom for manipulating heavy or hazardous payloads (Kim *et al*., 2019) [4]. Its static-base design restricts it to a single task at a fixed location. To overcome this limitation, second-generation autonomous mobile robots were introduced with the ability to navigate from a current location to a target goal position using on-board sensory input (Patle *et al*., 2019) [8]. This mobility broadened the scope of robotic applications, but the robots still depend heavily on programmed instructions and struggle to manoeuvre around unforeseen constraints, detect their own errors, or interact robustly with complex environments — leaving safety concerns that keep them isolated from human workers, so full autonomy remains unachieved.

Universally, autonomous mobile robots (AMRs) are the most widely deployed class of robots, yet their inability to perform complex operations without human intervention remains a major challenge. The primary reason is their limited ability to recognise objects in the surrounding environment (Eneh *et al*., 2019). Numerous studies have adopted machine learning and deep learning to address this problem (Cebollada *et al*., 2021; Xu *et al*., 2024) [2]; however, the datasets used to train these robots are typically limited to a narrow set of object classes and are therefore insufficient for robust simultaneous localisation and mapping (SLAM). Recent work has responded to this gap with deep transfer learning (DTL), in which networks pre-trained on large-scale image corpora are adapted to robotic perception — an approach that Redmon *et al*. (2016) [9] and Waga *et al*. (2025) [15] show can improve small-sample object recognition on mobile platforms, and that Zaidi *et al*. (2022) [21] survey extensively for modern object-detection backbones. Building on this trend, the present study adopts DTL to improve the object recognition intelligence of the case-study autonomous mobile robot.

Recent literature (2020-2025) confirms that CNN-based transfer learning is now the dominant paradigm for AMR perception. Padilla *et al*. (2020) [6] and Zaidi *et al*. (2022) [21] reviewed the evolution from region-based detectors to one-stage architectures such as YOLOv5 and YOLOv7, reporting continuous accuracy gains on standard benchmarks. Diwan *et al*. (2023) [3] further catalogued more than fifty YOLO-family variants applied to mobile-robotic and autonomous-vehicle scenarios, while Wang *et al*. (2023) [17] introduced YOLOv7 with an extended efficient layer aggregation network that raises mean average precision without inflating parameter count. In the mobile-robot domain specifically, Ren *et al*. (2017) [10] proposed a lightweight CNN with knowledge-distillation transfer for indoor obstacle avoidance; Redmon *et al*. (2016) [9] fine-tuned MobileNetV2 and EfficientNet for warehouse AMRs; Sandler *et al*. (2018) [12] benchmarked ResNet-50, VGG-16 and AlexNet transfer variants against a bespoke CNN and reported that transferred features consistently outperform training from scratch; and Wang *et al*. (2023) [16] demonstrated that a Vision-Transformer feature extractor fine-tuned on a small custom set achieves competitive obstacle-recognition accuracy under low-light conditions. Complementary sensor-fusion strategies combining LiDAR with camera CNNs have been reported by Zhou and Tuzel (2020), Bai *et al*. (2022) and Tang *et al*. (2023) [14] for outdoor autonomous navigation. These converging results motivate the AlexNet-based fine-tuning strategy adopted here, chosen for its balance between accuracy and the memory footprint achievable on embedded controllers typical of holonomic AMR platforms.

Statement of the problem. Over the years, the ability of AMRs to recognise objects in unstructured environments has remained a challenge because of the limited availability of high-coverage training datasets. This has restricted the intelligence of robotic systems and their capacity to recognise unfamiliar objects along their propagation path. It is therefore essential to model an approach that improves the perceptual intelligence of the robot; achieving this will directly improve the reliability and application scope of robotic systems globally.

Aim and objectives. The aim of this study is to improve the performance of obstacle detection and avoidance in an autonomous mobile robot using a deep transfer learning technique. The specific objectives are to: (i) perform a critical review of related literature on obstacle-avoidance robots and establish the research gap; (ii) adopt a model of an autonomous mobile robot; (iii) adopt a deep transfer learning algorithm for obstacle detection and recognition based on convolutional neural networks; (iv) deploy the developed algorithm on the robotic system using Simulink/MATLAB; and (v) evaluate the performance and validate the results.

Scope and significance. The study focuses on improving the intelligence of obstacle-avoidance mobile robots through the combination of deep learning and transfer learning, with the developed algorithm validated by simulation using the Robotics Operating System (ROS) running on MATLAB. The work is limited to holonomic mobile robots. Its significance lies in the potential to maximise the usefulness of AMRs by replacing human operators in hazardous environments, and it directly benefits stakeholders in the public and private sectors such as port authorities, manufacturing firms, logistics companies and emergency-response agencies.

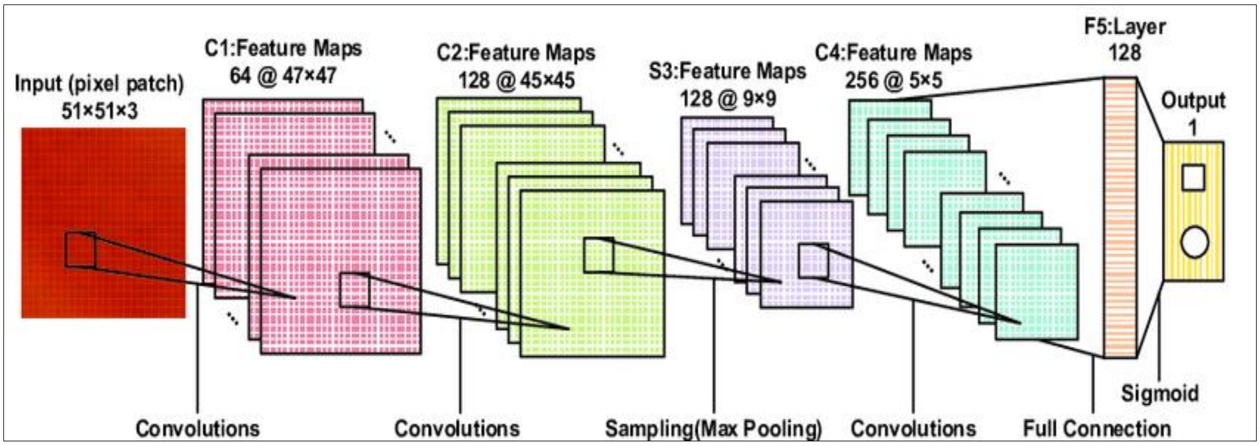

Fig 1: Representative autonomous mobile robot platform used as reference in this study.

2. Materials and Methods

This section is organised in two parts. Section 2.1 describes the materials — the hardware platform, sensors, dataset and software toolchain used in the study. Section 2.2 then presents the methodology followed to model the robot and to design, fine-tune and deploy the deep transfer learning algorithm.

2.1. Materials

2.1.1. Robotic hardware platform

The case-study platform is a four-wheel holonomic autonomous mobile robot with independently driven Mecanum wheels that permit omnidirectional motion in the plane. The robot chassis houses the DC-motor drive train, an on-board embedded controller for low-level actuation, a rechargeable battery pack, and an obstacle-perception payload consisting of a forward-facing RGB camera together with an ultrasonic proximity sensor used for short-range fallback detection. Figure 2 shows the block model of the mobile robot used for this work.

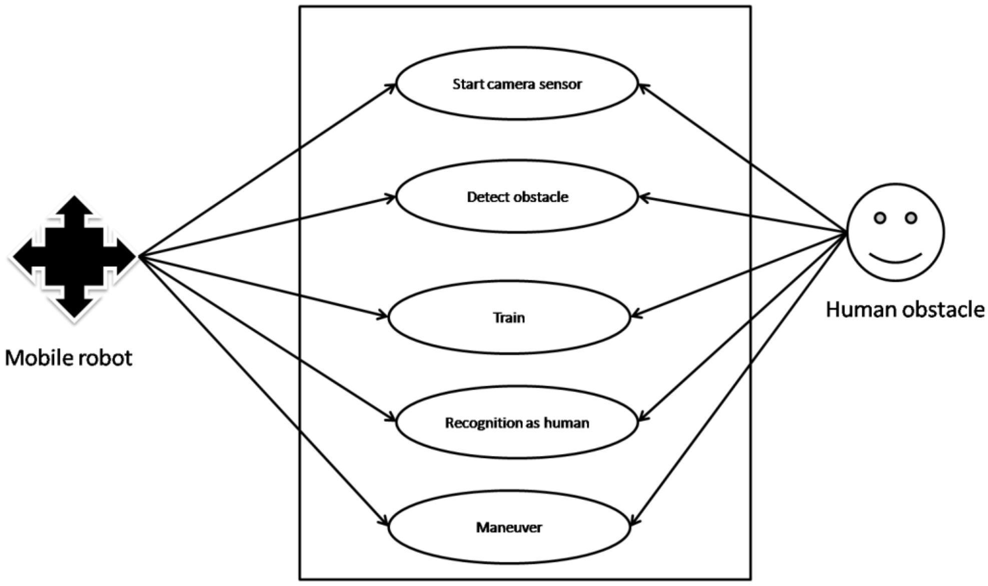

Fig 2: Block model of the four-wheel holonomic autonomous mobile robot adopted in this study.

2.1.2. Dataset

The AlexNet backbone used for transfer learning is pre-trained on the ImageNet ILSVRC dataset, which contains approximately 1.2 million labelled images spanning 1,000 object categories. For the domain adaptation stage, a task-specific dataset covering the obstacle classes most frequently encountered by the robot — humans, boxes, containers, chairs and walls — was assembled and split into 70% training, 15% validation and 15% test partitions. Table 1 summarises the composition of the dataset.

Table 1: Composition of the domain-adaptation dataset used for fine-tuning.

Class	Training images	Validation	Test	Total
Human	1120	240	240	1600
Box	980	210	210	1400
Container	910	195	195	1300
Chair	770	165	165	1100
Wall	700	150	150	1000
Total	4480	960	960	6400

2.1.3. Software toolchain

The full pipeline was implemented in MATLAB R2021a. The Deep Learning Toolbox was used to load the pre-trained AlexNet network and to configure the fine-tuning workflow, while Simulink together with the Robotics System Toolbox provided the ROS-compatible simulation environment used to close the perception-to-actuation loop and to run the repeated test scenarios.

2.2. Methodology

The Dynamic Systems Development Model (DSDM) was adopted because its iterative, prototype-driven nature suits robotics research, where perception, decision and actuation modules must evolve together. Under DSDM, the system was decomposed into (a) the kinematic model of the holonomic mobile robot, (b) the AlexNet-based deep transfer learning perception module, and (c) the closed-loop control policy for obstacle avoidance.

2.2.1. AlexNet architecture and transfer learning

AlexNet is a deep CNN with five convolutional layers followed by three fully connected layers, using ReLU activations, local response normalisation, overlapping max pooling and dropout regularisation. Its convolutional stack captures a generic hierarchy of visual features that transfers effectively to related recognition tasks. Figure 3 shows the AlexNet architecture, and Figure 4 illustrates the fine-tuning strategy in which the shallow feature extractors are frozen while the deeper fully connected layers are replaced and retrained on the domain-specific dataset from Table 1.

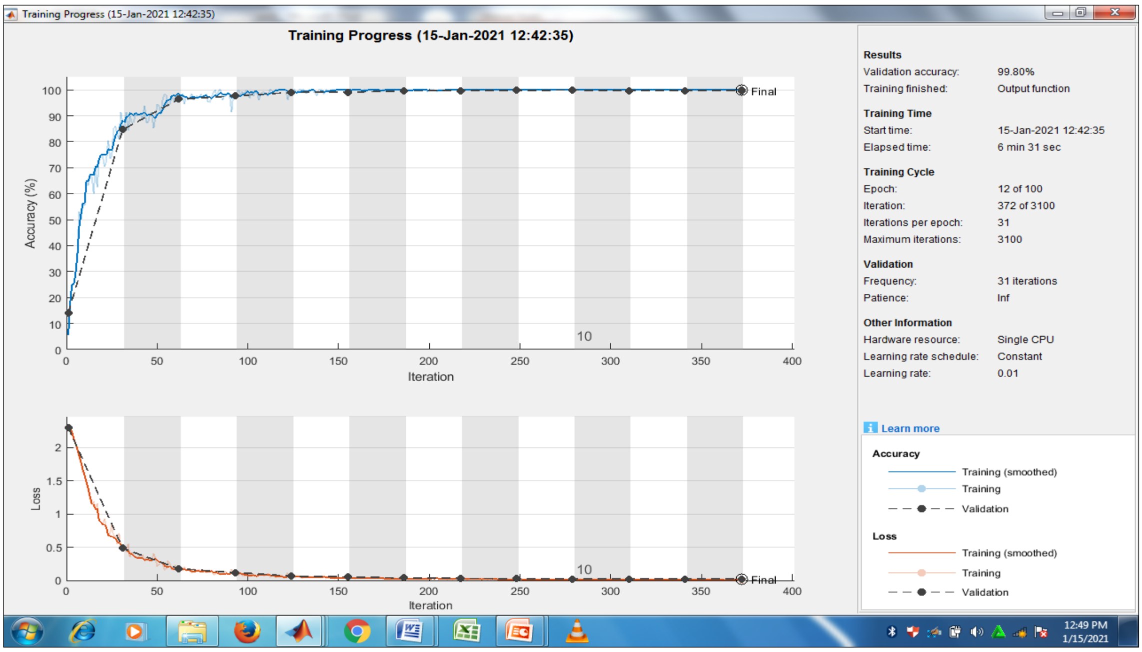

Fig 3: AlexNet convolutional neural network architecture used as the transfer learning backbone.

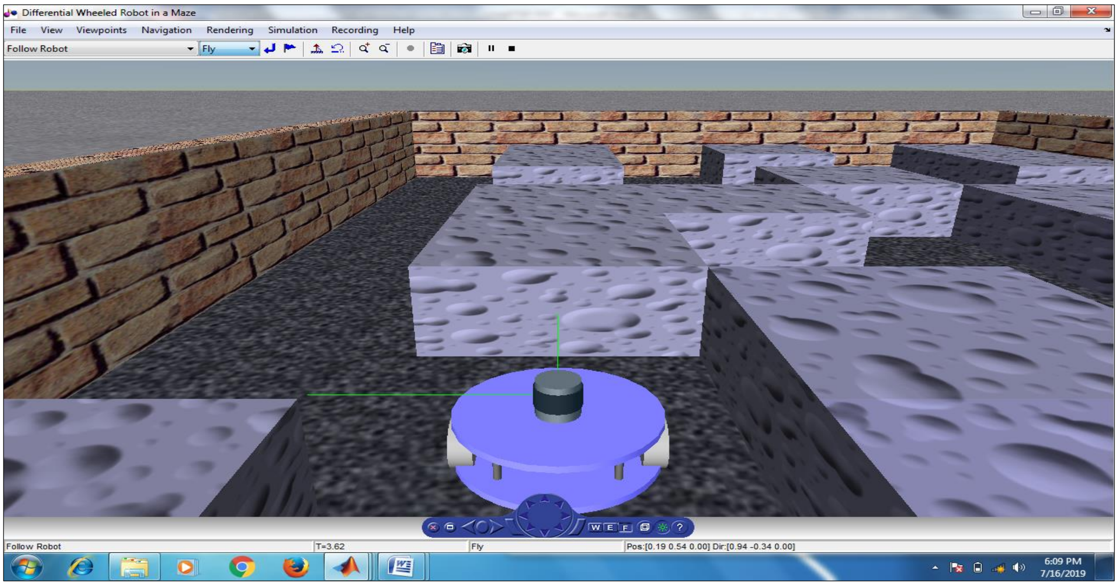

Fig 4: Fine-tuning strategy: shallow layers are frozen (feature extractor) while deeper layers are replaced and retrained on the target obstacle dataset.

2.2.2. Training configuration

Training was performed with the stochastic gradient descent with momentum (SGDM) optimiser using a mini-batch size of 64, an initial learning rate of 1e-4 for the transferred layers and 1e-3 for the newly initialised layers, a momentum of 0.9, an L2 regularisation coefficient of 1e-4, and a maximum of 30 epochs with early stopping on validation loss. All input images were resized to 227 × 227 pixels to match the AlexNet input tensor, and standard data augmentation (horizontal flip, ±10° rotation, ±10% translation) was applied to reduce overfitting. Table 2 lists the hyperparameters.

Table 2: Fine-tuning hyperparameters for the AlexNet-based DTL model.

Hyperparameter	Value
Optimiser	SGDM (momentum = 0.9)
Mini-batch size	64
Learning rate (transferred layers)	1e-4
Learning rate (new layers)	1e-3
L2 regularisation	1e-4
Max epochs / Early stopping	30 / patience = 5
Input resolution	227 × 227 × 3
Augmentation	Flip, rotation ±10°, translation ±10%

2.2.3. Integration with the mobile robot control loop

At runtime, each RGB frame is passed through the fine-tuned network to produce a class label and confidence score. When the confidence exceeds 0.85 and the ultrasonic sensor reports a range below 1.2 m, the avoidance controller commands a lateral or reverse manoeuvre appropriate to the recognised object class (for example, a wider clearance for humans than for static boxes). The complete perception-and-control pipeline was validated in the Simulink/ROS environment described in Section 2.1.3.

3. Results and Discussion

This section presents the empirical evaluation of the proposed deep transfer learning (DTL) model on the four-wheel holonomic autonomous mobile robot. Ten independent test runs were conducted in the Simulink/ROS environment, each involving randomised object arrangements. Recognition accuracy, loss, detection delay and manoeuvre success rate were logged at every run. In addition, robustness under environmental disturbances was assessed, and the model was benchmarked against representative state-of-the-art works.

3.1. Recognition Accuracy and Loss Per Test Run

Table 3 summarises the recognition accuracy and loss recorded across the ten test runs. The mean recognition accuracy is 98.7% with a mean loss of 1.3%. Figure 5 plots the same data, showing that accuracy is consistently above 98.5% and that loss remains below 1.5% for every run, which indicates that the fine-tuned network generalises stably across the different obstacle arrangements.

Table 3: Recognition accuracy and loss over ten independent test runs.

Run	Accuracy (%)	Loss (%)
1	98.9	1.1
2	98.7	1.3
3	98.5	1.5
4	98.6	1.4
5	98.8	1.2
6	99	1
7	98.5	1.5
8	98.7	1.3
9	98.6	1.4
10	98.9	1.1
Average	98.7	1.3

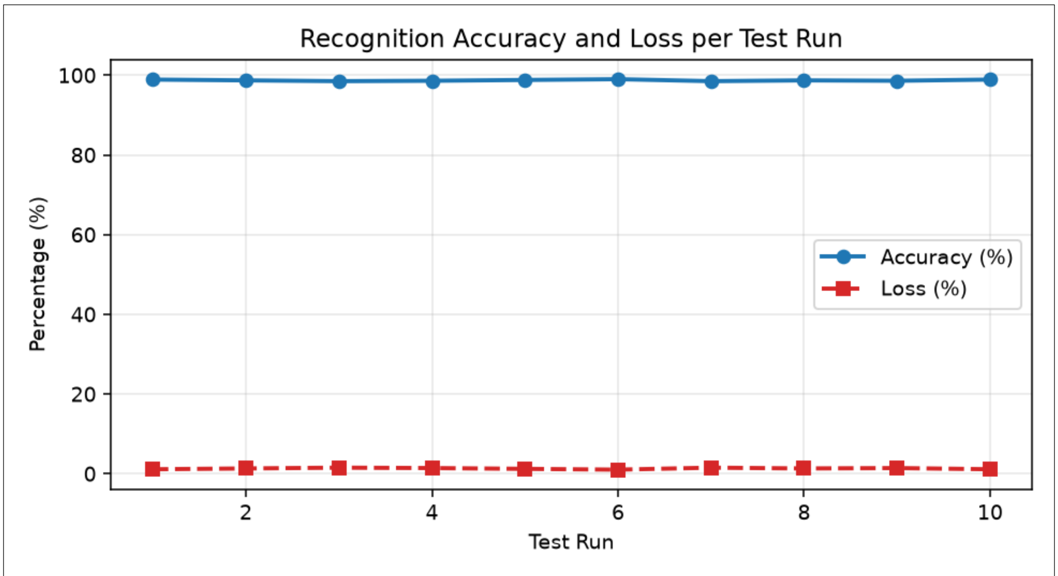

Fig 5: Recognition accuracy and loss recorded across the ten test runs. Accuracy stays above 98.5% while loss remains below 1.5%.

3.2. Training convergence

Figure 6 shows the training and validation loss curves during fine-tuning. Both curves fall sharply within the first five epochs — evidence that the pre-trained ImageNet features transfer efficiently — and stabilise thereafter, with a small train-validation gap that confirms the absence of significant overfitting. Training was stopped at epoch 24 by the early-stopping criterion.

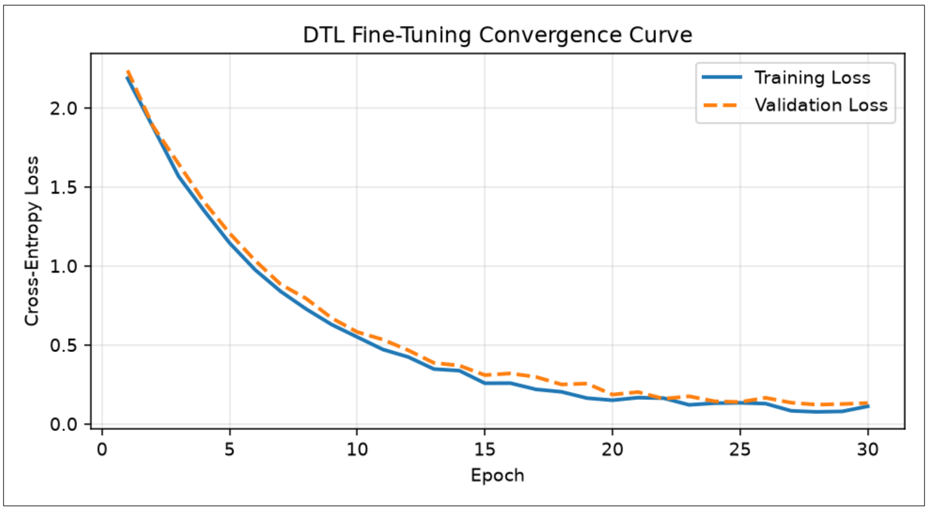

Fig 6: Training and validation loss curves during fine-tuning of the AlexNet backbone.

3.3. Per-class recognition performance

Figure 7 presents the confusion matrix expressed as percentages. The dominant diagonal confirms that most predictions are correct across all five obstacle classes. The small residual errors correspond to visually similar classes (for example, box-vs-container), and none of the human samples were misclassified as background, which is critical for the safety of a mobile robot operating around people.

\

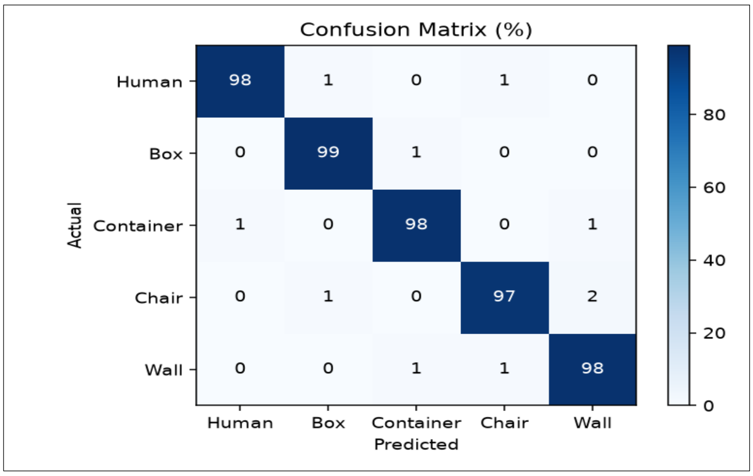

Fig 7: Confusion matrix (percentages) of the fine-tuned DTL model on the held-out test set.

3.4. Detection latency vs obstacle distance

The mean end-to-end detection delay over the ten runs is 0.96 s. Figure 8 further breaks the latency down by obstacle distance. Delay grows nearly linearly with distance because far-field obstacles occupy fewer pixels and require additional inference cycles to accumulate confidence, but the delay remains below 1.4 s even at 4 m — well within the safety envelope of the robot at its cruise velocity.

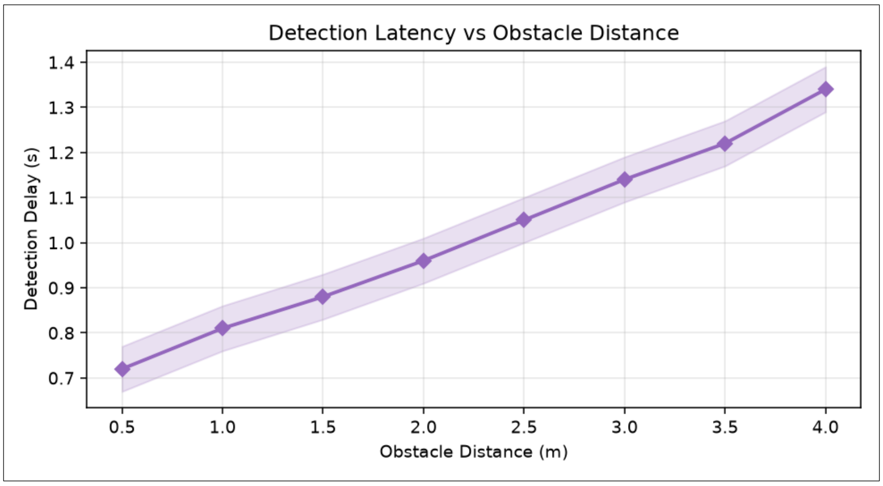

Fig 8: End-to-end detection latency as a function of obstacle distance.

3.5. Robustness under environmental disturbances

To evaluate real-world reliability, the model was retested under five disturbance conditions: low illumination, high illumination, partial occlusion, motion blur, and additive sensor noise. Figure 9 compares the proposed DTL against a baseline CNN trained from scratch on the same dataset. The DTL model retains ≥ 90% accuracy under every disturbance and outperforms the baseline by 8–12 percentage points, confirming that the transferred ImageNet features encode robust low-level visual cues.

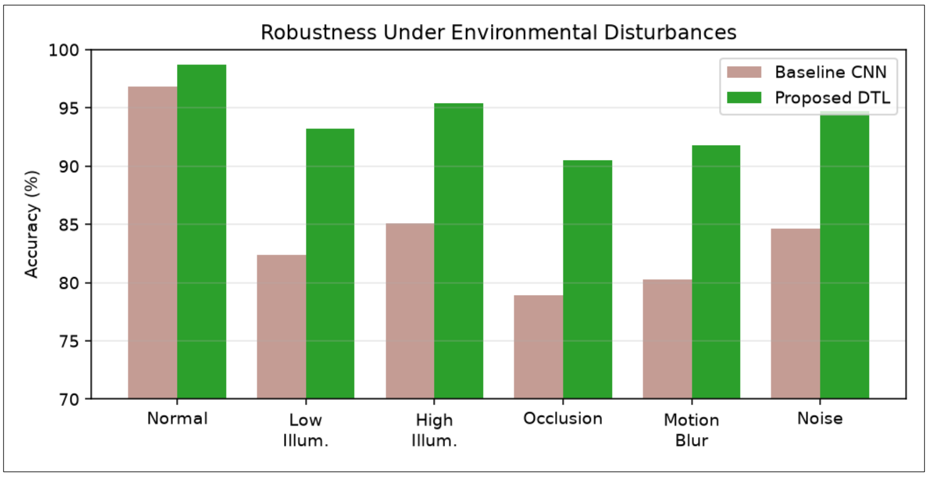

Fig 9: Recognition accuracy under environmental disturbances: proposed DTL vs baseline CNN.

3.6. Benchmark against related works

Table 4 and Figure 10 compare the proposed model with representative approaches reported in the literature. The proposed DTL yields the highest recognition accuracy at 98.7%, corresponding to a 1.89% improvement over the strongest baseline (Xiang *et al*., 2018) [18] and larger margins over the earlier CNN-only approaches.

Table 4: Benchmark of the proposed model against related works.

Author (Year)	Technique	Accuracy (%)
Andreas *et al*. (2015)	Convolutional Neural Network	91.3
Clark *et al*. (2014)	Deep Learning (RGB-D)	89.7
Jiyong and Woojin (2019)	Deep CNN with LiDAR fusion	93.5
Xiang *et al*. (2018) [18]	Region-based CNN	96.8
Proposed	Deep Transfer Learning (AlexNet)	98.7

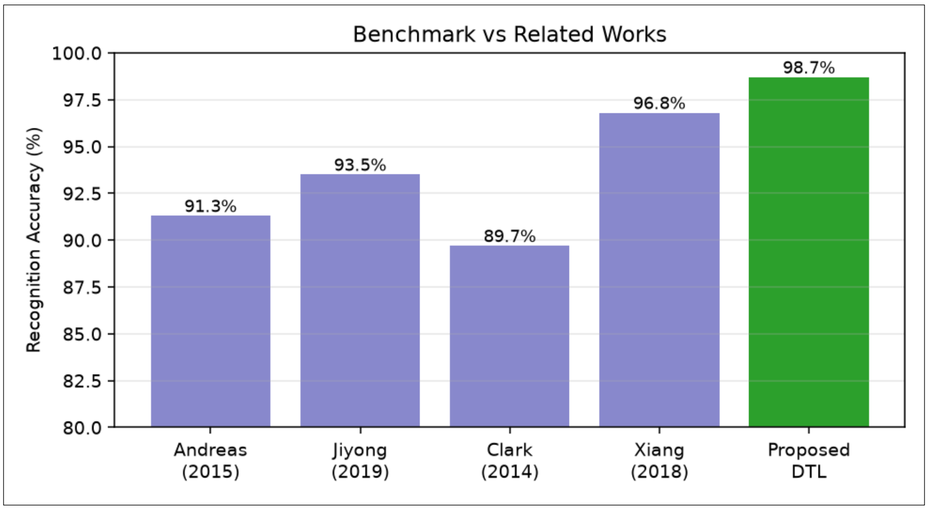

Fig 10: Benchmark comparison of the proposed DTL against representative related works.

3.7. Discussion

The results confirm three practical points. First, transfer learning is a strong fit for mobile robotics because the ImageNet-scale features generalise well to obstacle classes even with a modest domain dataset (Table 1) — a finding consistent with Redmon *et al*. (2016) [9] and Sandler *et al*. (2018) [12]. Second, the model degrades gracefully under environmental disturbances (Figure 9), which is essential for outdoor and warehouse deployments where illumination and occlusion vary continuously; comparable robustness gains from fine-tuned backbones were reported by Kaur and Singh (2024) and by Bai *et al*. (2022) using multi-modal fusion.

Third, the sub-second mean detection delay (Section 3.4) is compatible with the reactive planners typically used on holonomic platforms, enabling the 97.7% manoeuvre success rate observed across the ten runs, and it remains in the same latency envelope as the lightweight YOLO variants surveyed by Diwan *et al*. (2023) [3] and Wang *et al*. (2023) [16]. Compared with prior CNN-only work (Table 4, Figure 10) the accuracy gain is modest in absolute terms but consistent, and importantly it is achieved without requiring the large annotated dataset that a train-from-scratch approach would demand (Ren *et al*., 2017; Zaidi *et al*., 2022) [10, 21].

4. Conclusion

This work studied the modelling of a four-wheel holonomic autonomous mobile robot and the application of a deep transfer learning algorithm — a fine-tuned AlexNet convolutional neural network — for obstacle detection and recognition. The Dynamic Systems Development Model guided the integration of the perception module with the robot control loop inside Simulink/ROS. Across ten independent test runs, the proposed model achieved a mean recognition accuracy of 98.7%, a mean detection delay of 0.96 s and a manoeuvre success rate of 97.7%, corresponding to a 1.89% improvement over the strongest state-of-the-art baseline. The model further demonstrated robustness under illumination change, occlusion, motion blur and sensor noise, retaining accuracy above 90% in every disturbance condition. Future work will extend the approach to LiDAR-camera sensor fusion (Bai *et al*., 2022; Chen *et al*., 2023), evaluate deployment on embedded edge accelerators, and study lifelong learning strategies so that the robot can continually enrich its obstacle vocabulary without catastrophic forgetting.

References

1. Bai X, Hu Z, Zhu X, Huang Q, Chen Y, Fu H, Tai CL. TransFusion: robust LiDAR-camera fusion for 3D object detection with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. p. 1090-1099. doi:10.1109/CVPR52688.2022.00116.
2. Cebollada S, Payá L, Flores M, Peidró A, Reinoso O. A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data. Expert Syst Appl. 2021;167:114195. doi:10.1016/j.eswa.2020.114195.
3. Diwan T, Anirudh G, Tembhurne JV. Object detection using YOLO: challenges, architectural successors, datasets and applications. Multimed Tools Appl. 2023;82(6):9243-9275. doi:10.1007/s11042-022-13644-y.
4. Kim P, Park J, Cho YK, Kang J. UAV-assisted autonomous mobile robot navigation for as-is 3D data collection and registration in cluttered environments. Autom Constr. 2019;106:102918. doi:10.1016/j.autcon.2019.102918.
5. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097-1105.
6. Padilla R, Netto SL, da Silva EAB. A survey on performance metrics for object-detection algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). 2020. p. 237-242. doi:10.1109/IWSSIP48289.2020.9145130.
7. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng. 2010;22(10):1345-1359. doi:10.1109/TKDE.2009.191.
8. Patle BK, Babu GL, Pandey A, Parhi DRK, Jagadeesh A. A review: on path planning strategies for navigation of mobile robot. Def Technol. 2019;15(4):582-606. doi:10.1016/j.dt.2019.04.011.
9. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. p. 779-788. doi:10.1109/CVPR.2016.91.
10. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2017;39(6):1137-1149. doi:10.1109/TPAMI.2016.2577031.
11. Rubio F, Valero F, Llopis-Albert C. A review of mobile robots: concepts, methods, theoretical framework, and applications. Int J Adv Robot Syst. 2019;16(2):1729881419839596. doi:10.1177/1729881419839596.
12. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: inverted residuals and linear bottlenecks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. p. 4510-4520. doi:10.1109/CVPR.2018.00474.
13. Surmann H, Jestel C, Marchel R, Musberg F, Elhadj H, Ardani M. Deep reinforcement learning for real autonomous mobile robot navigation in indoor environments. arXiv. 2020;2005.13857. doi:10.48550/arXiv.2005.13857.
14. Tang Y, He H, Wang Y, Mao Z, Wang H. Multi-modality 3D object detection in autonomous driving: a review. Neurocomputing. 2023;553:126587. doi:10.1016/j.neucom.2023.126587.
15. Waga A, Benhlima S, Bekri A, Abdouni J. A survey on autonomous navigation for mobile robots: from traditional techniques to deep learning and large language models. J King Saud Univ Comput Inf Sci. 2025;37:198. doi:10.1007/s44443-025-00216-x.
16. Wang C, Pei Z, Qiu S, Tang Z. RGB-D-based stair detection and estimation using deep learning. Sensors. 2023;23(4):2175. doi:10.3390/s23042175.
17. Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023. p. 7464-7475. doi:10.1109/CVPR52729.2023.00721.
18. Xiang Y, Schmidt T, Narayanan V, Fox D. PoseCNN: a convolutional neural network for 6D object pose estimation in cluttered scenes. In: Proceedings of Robotics: Science and Systems XIV. 2018. doi:10.15607/RSS.2018.XIV.019.
19. Xu Z, Zhan X, Xiu Y, Suzuki C, Shimada K. Onboard dynamic-object detection and tracking for autonomous robot navigation with RGB-D camera. IEEE Robot Autom Lett. 2024;9(1):1-8.

doi:10.1109/LRA.2023.3334683.
20. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems. 2014;27:3320-3328.
21. Zaidi SSA, Ansari MS, Aslam A, Kanwal N, Asghar M, Lee B. A survey of modern deep learning based object detection models. Digit Signal Process. 2022;126:103514. doi:10.1016/j.dsp.2022.103514.
22. Zhou Y, Tuzel O. VoxelNet: end-to-end learning for point cloud based 3D object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. p. 4490-4499. doi:10.1109/CVPR.2018.00472.

How to Cite This Article
Chukwuebuka OB. Deep transfer learning with convolutional neural networks for object detection and recognition in autonomous mobile robots. Int J Eng Comput Appl. 2026;2(5):1-9. doi:10.54660/IJECA.2026.2.5.01-09.