UJEAS

UNIZIK JOURNAL OF ENGINEERING AND APPLIED SCIENCES

UNIZIK Journal of Engineering and Applied Sciences 6(5), December (2026), 4114-4121
Journal homepage: https://journals.unizik.edu.ng/index.php/ujeas
PRINT ISSN: 2992-4383 || ONLINE ISSN: 2992-4391

Comparative Pre-Trained Models for Obstacle Avoidance Application in Indoor Environment

[1]Ebere Uzoka Chidi, [2]Udeh Chukwuma Callistus

[1]African Centre of Excellence for Sustainable Power and Energy Development, Department of Control and Instrumentation Engineering, University of Nigeria Nsukka, Enugu, Nigeria

[2]Department of Computer Science, Faculty of Applied and Natural Science, Enugu State University of Science and Technology (ESUT), Agbani, Nigeria

Email: [1]steverolans@gmail.com, [2]chukwuma.udeh@esut.edu.ng

Abstract

An obstacle is an object that lies along the intended path of a moving person or device. Collision with such objects can cause injury, particularly for people with visual impairment in indoor spaces. This paper presents a comparative study of pre-trained models for indoor obstacle avoidance. Previous studies often treat every detected object as an obstacle and therefore generate false alarms. This paper separates objects of interest from true obstacles by combining transfer learning with a Distance Estimation Algorithm (DEA). MobileNet-V2 with a Single Shot MultiBox Detector (SSD) head, DenseNet-121 used as an image-level classification baseline, and You Only Look Once version 8 nano (YOLOv8n) were fine-tuned on a working set of 1,010 labelled indoor images of chairs and tables. The initial capture set comprised 59 photographs (45 chairs and 14 tables), and geometric augmentation was used to expand the training material. On the reported test split, YOLOv8n achieved 58.8% mAP@0.50:0.95 with 3.2 million parameters and a reported desktop inference time of 4.5ms, compared with 25.1% mAP@0.50:0.95 for MobileNet-V2+SSD. The selected YOLOv8n detector was integrated with a fully defined Distance Estimation Algorithm and a 2.5 m indoor obstacle threshold. Comparative results identified YOLOv8n as the more suitable of the two object detectors for integration with the DEA. The obstacle decision threshold was revised from 20m to 2.5m for indoor navigation, and the DEA was evaluated by comparing camera-estimated distances with tape-measured distances. Experimental checks in three indoor settings were then used to assess obstacle detection and classification. In nine trials comparing camera estimates with tape measurements, the mean absolute error for the eight objects measured within 2.5m was 0.09m. Seven of those eight objects triggered an alert; one chair measured at 2.40 m was estimated at 2.52 m and was missed. Tests in three indoor settings illustrate the method, but the small number of original photographs and the missed alert near the threshold limit the conclusions. Further testing with more independent images and portable hardware is needed.

Keywords: Computer vision; object of interest; transfer learning; YOLOv8n; visual impairment; distance estimation algorithm

1. Introduction

Independent walking requires more than intact limbs. Vision supplies object detection, recognition, spatial layout and the timing of avoidance decisions. Visual impairment therefore reduces everyday mobility even when gait is otherwise normal. Recent global estimates place the number of people who are completely blind at about 43.3 million, with hundreds of millions living with some form of visual impairment (World Health Organization, 2024; Zou and Zhou, 2023). Conventional aids such as the long cane, guide dogs and many wearable buzzers remain limited in how early and how reliably they describe the identity and range of an indoor obstacle (Chidi et al., 2025). Computer vision systems that detect and name objects have therefore been proposed as complementary tools. Transfer learning reduces the amount of labelled data needed to adapt a pre-trained network to a new indoor scene (Chidi et el., 2024). One-stage detectors, and the YOLO family in particular, are widely preferred when inference must stay near real time (Oluwaseyi et al. 2023). Several authors have added attention blocks or specialised necks to YOLO variants in order to raise detection accuracy (Rya and Chung, 2021). Those accuracy gains do not, by themselves, decide whether a detected chair is an obstacle on the user's path or an object sitting safely to one side. Figure 1 illustrates the resulting reliability gap.

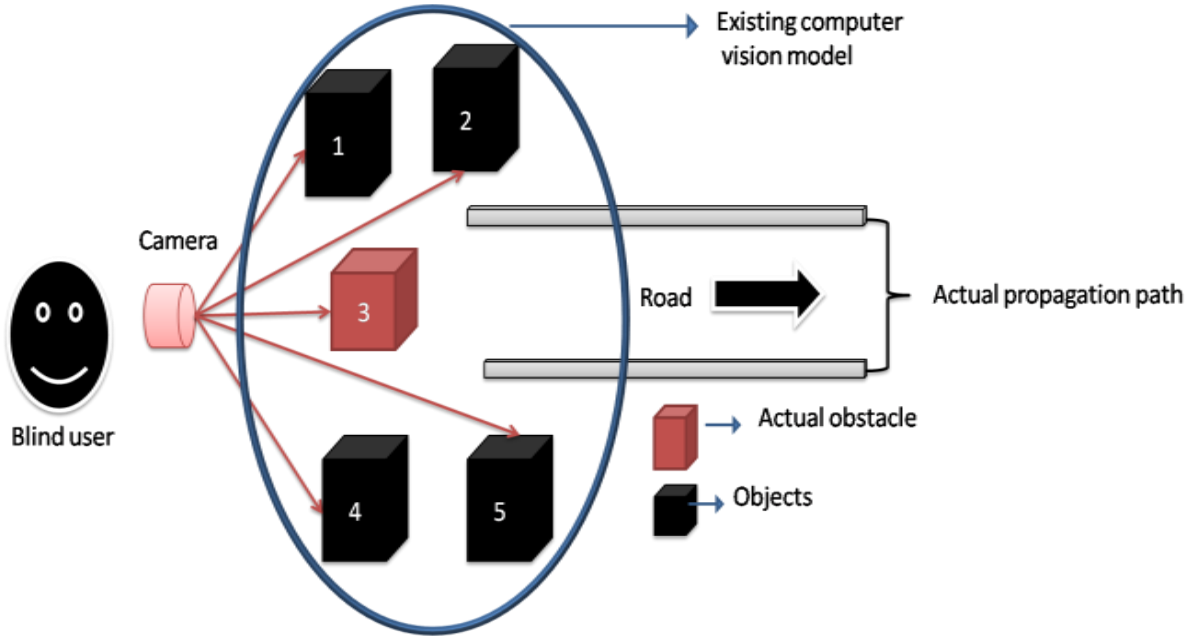

Figure 1: Existing object-only pipeline that does not separate objects from path obstacles.

The novelty of this study is the joint use of a lightweight one-stage detector (YOLOv8n) and a monocular Distance Estimation Algorithm that converts bounding-box geometry and camera focal length into a metric range. Only objects that both belong to the trained indoor classes and lie closer than a room-scale threshold are labelled as obstacles. The comparison with MobileNet-V2+SSD (a true detector) and DenseNet-121 (an image-level classifier reported only as a presence/absence baseline) makes the role of localisation explicit. The work therefore does not treat "object detected" and "obstacle present" as the same event. The experimental setting is the African Centre of Excellence for Sustainable Power and Energy Development (ACE-SPED) building, University of Nigeria, Nsukka, together with two further indoor rooms used for validation. Figure 2 summarises the proposed pipeline.

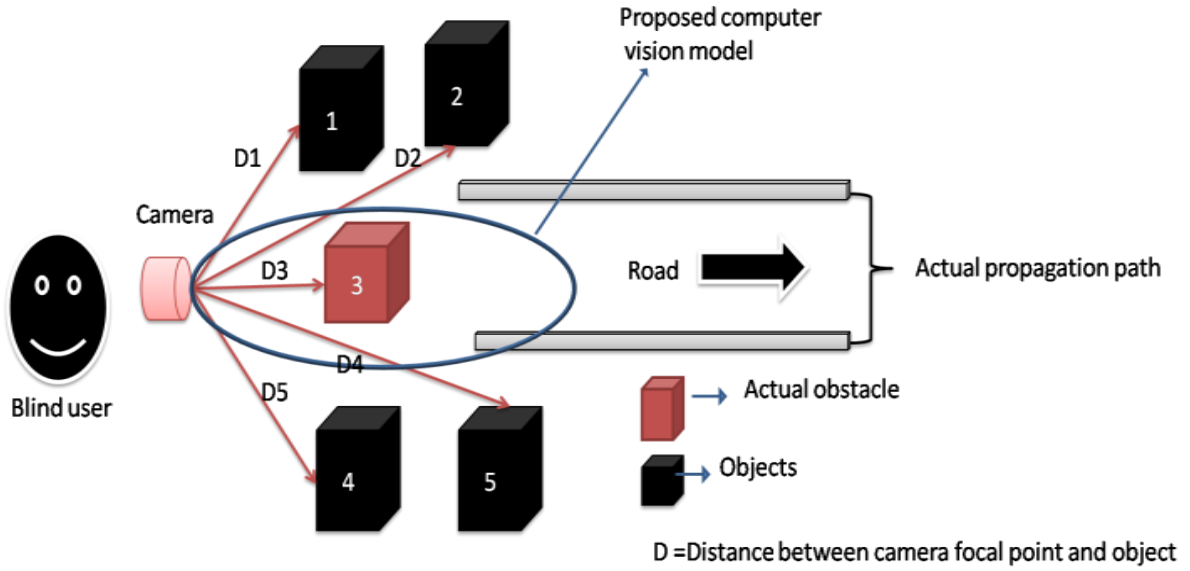

Figure 2: Proposed pipeline that couples detection with distance-based obstacle classification.

A clear gap remains in the indoor navigation literature. Studies that raise YOLO accuracy still report object labels without a calibrated indoor range test, while monocular ranging papers rarely couple the range estimate to a detector that has been compared against other pre-trained families on the same indoor furniture set (Khow wt al., 2024). This paper closes that gap by comparing three pre-trained families on the same chair–table corpus, attaching a documented DEA with fully defined variables to the winning detector, replacing an unrealistic 20 m gate with a 2.5 m indoor threshold, and reporting tape-measured versus estimated distances together with a short discussion of edge-device inference. These four elements constitute the innovative contribution relative to the earlier YOLO-V8n navigation model in (Chidi et al., 2025).

The contributions of the paper are as follows:

i. To extend the work in (Chidi et al., 2025) by including MobileNet-V2+SSD as a second object detector, while retaining DenseNet-121 only as a non-localising classification baseline.
ii. To integrate a clearly defined Distance Estimation Algorithm with the selected detector and classify obstacles using a 2.5 m indoor threshold.
iii. To validate the pipeline in three indoor environments, compare estimated and measured distances, and discuss practical deployment on portable edge hardware.

2. METHODOLOGY

Data collection targeted two indoor object classes that frequently block walking paths: chairs and tables. Images were captured with an iPhone X inside the African Centre of Excellence for Sustainable Power and Energy Development (ACE-SPED). The original capture set contained 45 chair photographs and 14 table photographs, giving 59 source images. Because this number of independent source images is limited, geometric augmentation was applied in Roboflow using rotations of 30°, 45° and 90° and left-right flips. The resulting working set contained 1,010 labelled images, comprising 520 chair images and 490 table images. Bounding boxes and class labels were prepared in the same Roboflow environment. Augmentation increases viewpoint variation but does not create genuinely new rooms or independent scenes; therefore, the small number of source photographs remains a limitation. To reduce the risk of interpreting memorisation as generalisation, practical validation was carried out in three indoor settings described in Section 4.2, and conclusions are restricted to the two trained furniture classes and the tested indoor conditions. YOLOv8n, MobileNet-V2+SSD and DenseNet-121 were fine-tuned on this set. DenseNet-121 is an image classifier and cannot output boxes; it is retained only as a multilabel presence/absence baseline so that the cost of omitting localisation is visible. The two true detectors are YOLOv8n and MobileNet-V2+SSD. The winning detector was then combined with the DEA described in Section 3.

2.1 YOLOv8n

YOLOv8n is the nano member of the YOLOv8 family. It comprises an input stem, a backbone of convolutional and C2f blocks with spatial pyramid pooling (SPPF), a neck that concatenates multi-scale maps, and a decoupled head that predicts class, box and distribution-focal terms. Figure 3 shows the layout used in this study.

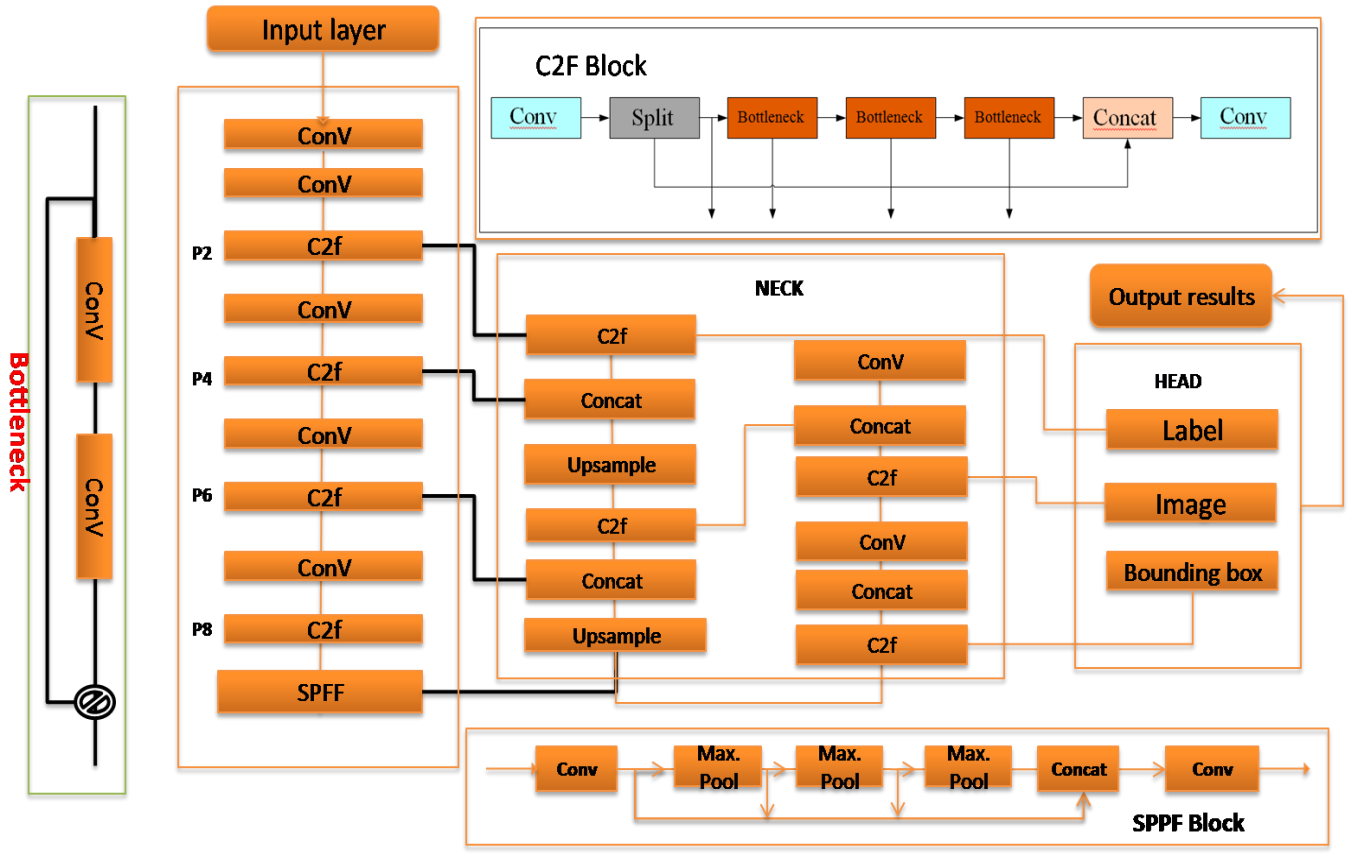

Figure 3: YOLOv8n architecture (Chidi et al., 2025).

2.2 MobileNet-V2 + SSD

MobileNet-V2 combined with a Single Shot MultiBox Detector head is a one-stage detector designed for modest compute budgets. The backbone uses inverted residual blocks with depth-wise and point-wise convolutions. The

SSD head predicts boxes on multi-scale feature maps. Input resolution in this work follows the common 300 × 300 SSD setting. Figure 4 outlines the stack.

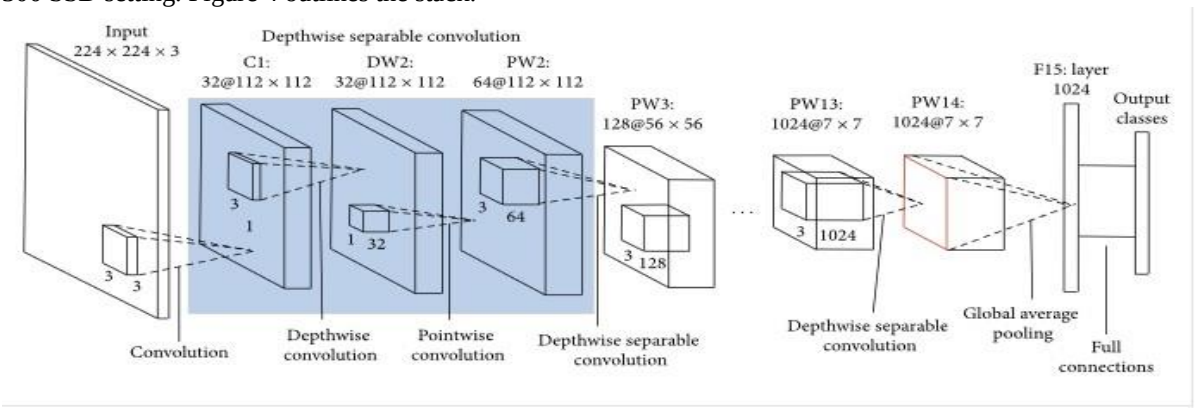

Figure 4: MobileNet-V2 + SSD architecture (Chidi et al., 2024).

2.3 DenseNet-121 as a classification baseline

DenseNet-121 is a densely connected convolutional classifier. Feature maps are reused across layers, which eases gradient flow and limits the number of parameters (Chidi et al., 2024). Because the network produces class scores for an entire image and not bounding boxes, it is not a drop-in rival to YOLOv8n. It is reported here only to show how well a strong classifier can answer the weaker question "is a chair or table present in this frame?" The comparison in Table 4 therefore separates detector metrics (mAP with localisation) from classifier metrics (multilabel mAP without boxes).

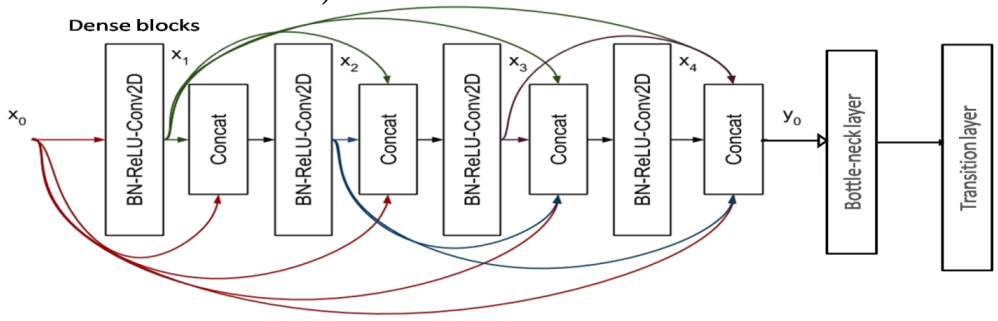

Figure 5: Block diagram of DenseNet-121.

2.4 Performance metrics

Precision, recall, F1-score and mean Average Precision are defined as follows (Ezeaku-Ezeme & Omankwu, 2024).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F}_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

$$\text{mAP} = \frac{1}{\text{N}} \sum \text{AP}_\text{k} \quad (6)$$

Where AP_k is the area under the precision-recall curve for class k. TP, FP and FN are true positives, false positives and false negatives at a chosen Intersection-over-Union (IoU) threshold.

3. DISTANCE ESTIMATION ALGORITHM (DEA)

The DEA converts a detected box and known camera geometry into an approximate range. All symbols used below are defined in the text so that the calculation can be reproduced.

Let F be the camera focal length in pixels. Let (c_x, c_y) be the principal point of the image (approximately the image centre for a well-calibrated phone camera). Let (x_1, y_1) be the top-left corner and (x_2, y_2) the bottom-right corner of the predicted box, measured in pixels. The box width and height in pixels are

$$\text{W}_\text{px} = \text{x}_2 - \text{x}_1, \quad \text{H}_\text{px} = \text{y}_2 - \text{y}_1 \quad (7)$$

Let W_{real} and H_{real} denote the measured physical width and height, in metres, of the reference object or object dimension used for calibration. The same physical dimension used during focal-length calibration must be used consistently during range estimation. The monocular range D in metres is

$$\text{D} = \frac{\text{W}_\text{real} \times \text{F}}{\text{W}_\text{px}} \quad (8)$$

When width is the more stable dimension, Equation (2) is used. When height is more stable, $D = \frac{H_{real} \times F}{H_{px}}$. The focal length in pixels is obtained from a one-time calibration with an object of known size at a known distance D_{cal}:

$$F = \frac{W_{px,cal} \times D_{cal}}{W_{real}} \quad (9)$$

An object is labelled an obstacle only if two conditions hold: (i) the detector assigns it to chair or table with confidence above 0.50, and (ii) the estimated range D is less than or equal to D_{th}. For indoor walking the threshold is D_{th} = 2.5 m. Distances greater than 2.5 m are treated as scene objects that do not yet require an avoidance cue. Among all boxes that satisfy both tests, the nearest object is reported first to the user.

4. RESULTS AND DISCUSSION

YOLOv8n was trained with the standard box, classification and distribution-focal losses. Table 1 records selected epochs.

Table 1: YOLOv8n training log (selected epochs)

Epoch	Train box	Train cls	Train dfl	Val box	Val cls	Val dfl	LR	Val mAP 0.5:0.95
1	2.35	2.51	1.92	2.48	2.62	2.01	1e-4	0.084
10	1.98	2.04	1.62	2.12	2.15	1.69	8e-4	0.241
30	1.62	1.52	1.31	1.76	1.63	1.38	7e-4	0.337
60	1.38	1.18	1.08	1.52	1.29	1.15	5e-4	0.382
100	1.21	0.97	0.93	1.36	1.09	1.01	3e-4	0.405
150	1.12	0.85	0.83	1.28	0.97	0.92	2e-4	0.419
200	1.07	0.78	0.77	1.24	0.89	0.86	1e-4	0.425
250	1.05	0.75	0.74	1.22	0.86	0.84	5e-5	0.427
300	1.04	0.73	0.73	1.21	0.85	0.83	1e-5	0.588

At epoch 300 the validation box, classification and DFL losses had flattened. Reported detection scores for YOLOv8n on the held-out split were precision 0.98, recall 0.98, mAP@0.50 = 0.97 and mAP@0.50:0.95 = 0.588.

Table 2: DenseNet-121 classification training (image-level baseline)

Epoch	Train BCE loss	Train top-1 acc.	Val loss	Val multilabel (mAP)
1	0.682	0.624	0.715	0.412
10	0.251	0.862	0.312	0.708
20	0.164	0.921	0.253	0.746
30	0.141	0.938	0.247	0.751

Table 3: MobileNet-V2 + SSD detector training

Epoch	Train total	Train loc	Train conf	Val total	Val mAP 0.5:0.95
1	4.82	2.21	2.61	5.03	0.039
20	2.51	1.14	1.37	2.74	0.204
80	2.04	0.87	1.17	2.32	0.248
200	1.92	0.79	1.13	2.24	0.251

DenseNet-121 reached a multilabel mAP of 0.751 at epoch 30. That figure describes presence/absence only. MobileNet-V2+SSD, which does localise objects, reached only 0.251 mAP@0.50:0.95. The shallow 300 × 300 SSD head and the modest number of small furniture instances explain the weaker detector score. YOLOv8n, trained at 640 × 640, reached 0.588 mAP@0.50:0.95 with 3.2 million parameters.

1.1. 4.1 Comparative results

Table 4: Comparison of the three pre-trained families

Model	Role	Parameters (M)	Inference (ms)	Score
DenseNet-121	Classifier only	7.0	4.2	75.1% multilabel mAP
YOLOv8n	Detector	3.2	4.5	58.8% mAP@0.5:0.95
MobileNet-V2+SSD	Detector	14.2	5.1	25.1% mAP@0.5:0.95

Table 4 must be read with the role column in mind. DenseNet-121 is not a detector; its higher percentage is an image-level score and cannot be ranked directly against mAP of a box predictor. Among the two detectors, YOLOv8n is clearly stronger and more parameter-efficient than MobileNet-V2+SSD. This ranking agrees with recent one-stage detector surveys that place YOLO nano models ahead of older SSD heads on small indoor objects [5, 10, 12]. YOLOv8n was therefore selected for DEA integration.

4.2 Integration with the DEA and indoor tests

The trained YOLOv8n head was wrapped with the ranging equations (1)–(2). Three indoor sites were used: (A) an ACE-SPED corridor bay, (B) an ACE-SPED laboratory bay, and (C) a separate indoor office. Figure 6 shows the sites; Figure 7 shows detections with range tags.

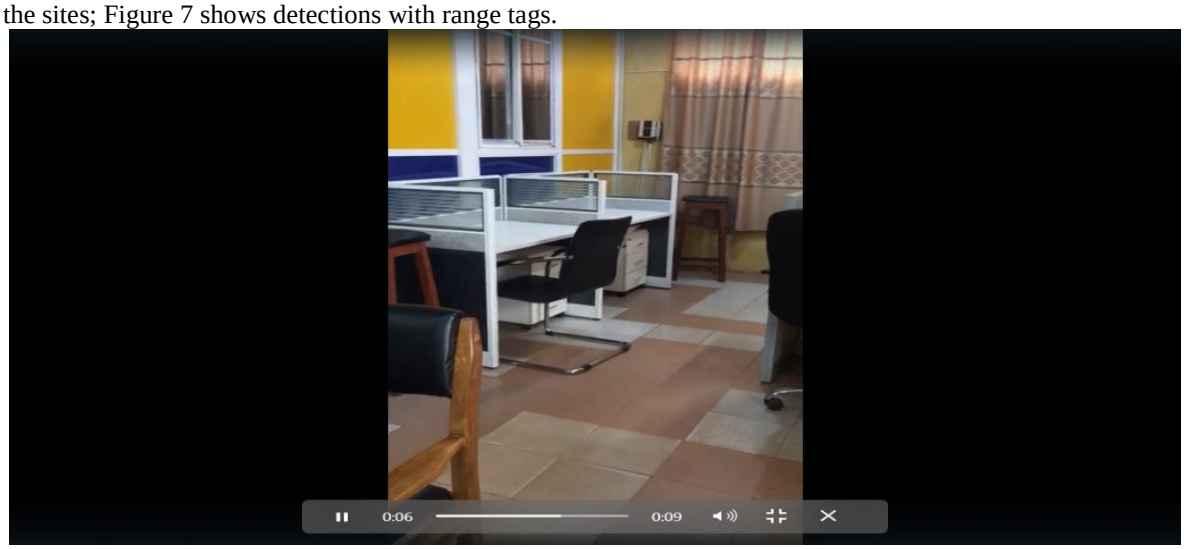

Figure 6: Experimental indoor sites (a) ACE-SPED bay, (b) ACE-SPED laboratory, (c) indoor office.

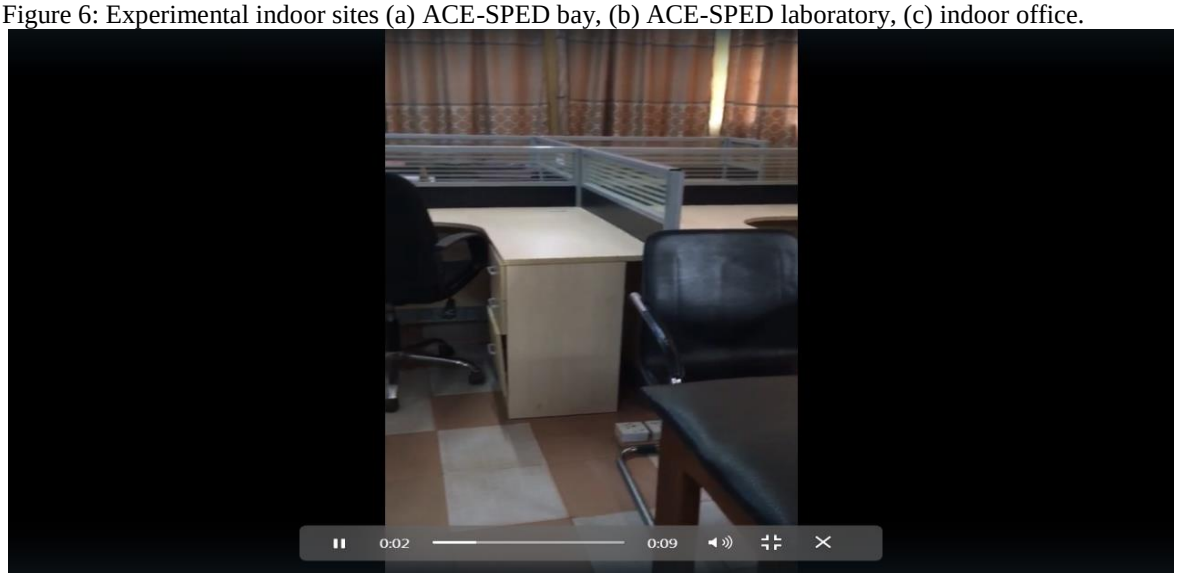

Figure 7: Practical validation. Detected chairs and tables are tagged with class and estimated range; objects inside 2.5 m are treated as obstacles.

To test whether the DEA returns usable numbers, a steel tape was used as ground truth at nine stand-off distances. Table 5 reports the comparison.

Table 5: Camera-estimated range versus tape-measured range (indoor, 2.5 m gate)

Trial	Object	Tape (m)	DEA estimate (m)	Absolute error (m)
1	Chair	0.80	0.86	0.06
2	Chair	1.20	1.27	0.07
3	Chair	1.80	1.71	0.09
4	Chair	2.40	2.52	0.12
5	Table	0.90	0.97	0.07
6	Table	1.50	1.41	0.09
7	Table	2.00	2.11	0.11
8	Table	2.50	2.38	0.12
9	Chair	3.20	3.41	0.21

Table 5 presents nine trials. The mean absolute error for trials 1–8, measured within the 2.5 m working range, was 0.09 m. Seven of these eight trials triggered an obstacle alert. The chair measured at 2.40 m was estimated at 2.52 m and was therefore missed by the 2.5 m alert rule. In trial 9, the chair measured at 3.20 m was correctly left outside the alert range. The larger distance error at 3.20 m is consistent with Equation (1), since a smaller image box makes the estimate more sensitive to changes in box width.

4.3 Edge hardware considerations

A navigation aid for a visually impaired user must ultimately operate on portable hardware. In the present study, YOLOv8n contains about 3.2 million parameters and the reported desktop inference time was 4.5 ms under the study's test configuration. This desktop result should not be interpreted as a measured Raspberry Pi, Jetson, or smartphone benchmark because those devices were not directly benchmarked in this experiment. Nevertheless, recent studies demonstrate that lightweight object detectors, including YOLOv8-family models, can be deployed on edge platforms. Lema et al. (2024) showed that object-detection performance on edge devices depends strongly on the device, input resolution and inference framework, while Gu et al. (2024) demonstrated practical YOLOv8-based deployment on an NVIDIA Jetson Orin Nano. For the present pipeline, practical deployment would therefore require exporting the trained detector to an edge-compatible format such as ONNX, TensorRT or TensorFlow Lite, measuring end-to-end latency on the selected device, and evaluating power consumption, camera acquisition, DEA computation, and audio or haptic feedback together. Raspberry Pi, NVIDIA Jetson and smartphones are feasible candidate platforms, but device-specific frame rate and battery life remain subjects for direct experimental evaluation.

Relative to the parent study (Chidi et al., 2025), the present work adds MobileNet-V2+SSD as a second detector, treats DenseNet-121 explicitly as a non-localising classifier baseline, uses a realistic indoor distance threshold, and includes a measured-versus-estimated ranging check. On the present dataset, YOLOv8n achieved an mAP@0.50:0.95 of 0.588 compared with 0.251 for MobileNet-V2+SSD. This comparison is valid between the two detectors because both localise objects, whereas the DenseNet-121 multilabel score answers a different classification task and is therefore not ranked directly against detector mAP. The DEA component further extends object detection into obstacle classification by requiring a detected object to fall within the 2.5 m distance gate before an avoidance cue is generated.

5. CONCLUSION

This study compared pre-trained models for detecting chairs and tables in indoor spaces. YOLOv8n and MobileNet-V2 with a Single Shot MultiBox Detector were evaluated as object detectors, while DenseNet-121 was used only as an image-level classification baseline. On the reported test split, YOLOv8n achieved 58.8% mean average precision at intersection-over-union thresholds from 0.50 to 0.95, compared with 25.1% for MobileNet-V2 with SSD. YOLOv8n had 3.2 million parameters and a reported desktop inference time of 4.5 ms. We combined YOLOv8n with a Distance Estimation Algorithm and used 2.5 m as the indoor alert threshold. Table 5 contains nine distance trials. For the eight trials measured within 2.5 m, the mean absolute error was 0.09 m. Seven of those eight trials triggered an alert. The chair measured at 2.40m was estimated at 2.52m and was missed because its estimate exceeded the threshold. The chair measured at 3.20m remained outside the alert range, as intended. These results show that distance estimation can help distinguish a nearby detected object from one farther away, but errors close to the threshold can affect the alert decision. The study is limited by its 59 original photographs, two furniture classes and three indoor test settings. The reported speed was measured on a desktop computer; portable hardware was not tested. Future work should use more independent images from unfamiliar buildings, include other obstacle types, improve calibration near the alert threshold and test the complete system on a portable device.

6. ACKNOWLEDGEMENT

The authors thank the African Centre of Excellence for Sustainable Power and Energy Development (ACE-SPED), University of Nigeria, Nsukka, for access to the indoor testbed.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI ChatGPT-5 in order to assist with language editing and grammar check. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

7. REFERENCES

Chidi, E. U., Anoliefo, E., Udanor, C., Chijindu, A. T., & Nwobodo, L. O. 2025. A blind navigation guide model for obstacle avoidance using distance vision estimation based YOLO-V8n. *Journal of the Nigerian Society of Physical Sciences, 7*(1), Article 2292. https://doi.org/10.46481/jnsps.2025.2292

Chidi, E. U., Udanor, C. N., & Anoliefo, E. 2024. *Exploring the depths of visual understanding: A comprehensive review on real-time object of interest detection techniques* [Preprint]. Preprints.org.

Ezeaku-Ezeme, C. N., & Omankwu, O. C. B. 2024. Early prediction of dementia using machine learning. *Caritas Journal of Physical and Life Sciences, 3*(2). https://caritasuniversityjournals.org/cjpls/article/view/115

Gu, Z., He, D., Huang, J., Chen, J., Wu, X., Huang, B., Dong, T., Yang, Q., & Li, H. 2024. Simultaneous detection of fruits and fruiting stems in mango using improved YOLOv8 model deployed by edge device. *Computers and Electronics in Agriculture, 228*, Article 109512. https://doi.org/10.1016/j.compag.2024.109512

Khow, Z., Tan, Y., Abdul, K. H., & Rashid, H. 2024. Improved YOLOv8 model for a comprehensive approach to object detection and distance estimation. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3396224

Lema, D. G., Usamentiaga, R., & García, D. F. 2024. Quantitative comparison and performance evaluation of deep learning-based object detection models on edge computing devices. *Integration, 95*, Article 102127. https://doi.org/10.1016/j.vlsi.2023.102127

Li, Y., Li, S., Du, H., Chen, L., Zhang, D., & Li, Y. 2021. YOLO-ACN: Focusing on small target and occluded object detection. *IEEE Access*. https://doi.org/10.1109/ACCESS.2020.3046515

Oluwaseyi, E., Martins, E., & Abraham, E. 2023. A comparative analysis of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics, 2*, 1-12. https://doi.org/10.33736/jcsi.5070.2023

Ryu, S., & Chung, K. 2021. Detection model of occluded object based on YOLO using hard-example mining and augmentation policy optimization. *Applied Sciences, 11*, Article 7093. https://doi.org/10.3390/app11157093

World Health Organization. 2024, February 10. *Blindness and vision impairment*. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment (Accessed, 20/4/2026)

Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y., & Wang, X. 2022. *YOLO-FaceV2: A scale and occlusion aware face detector* [Preprint]. arXiv. https://arxiv.org/abs/2208.02019

Zhao, Y., & Geng, S. 2021. Face occlusion detection algorithm based on YOLOv5. *Journal of Physics: Conference Series, 2031*, Article 012053. https://doi.org/10.1088/1742-6596/2031/1/012053

Zou, X., & Zhou, Y. 2023. Spatial cognition of the visually impaired: A case study in a familiar environment. *International Journal of Environmental Research and Public Health, 20*(3), Article 1753. https://doi.org/10.3390/ijerph20031753