AOI: 10.100.234513.0213
13th international conference on Microelectronics, Circuits and Systems, Micro2026

Real-Time Assistive Navigation Using an Edge-Optimized Transformer-Based Detection Framework

Potipireddi Kumari, Mukku Pavan Kumar

Department of Electronics & Communication Engineering
Lendi Institute of Engineering and Technology (A), Vizaganagaram, India
Corresponding author: pkumari.004@gmail.com

ABSTRACT

Reliable perception capabilities are essential for assistive navigation systems for the visually impaired, which must operate within the constraints of computing and energy limits of edge devices. On low-power embedded platforms, the proposed study aims to implement an object identification framework that is based on edge optimization of transformers. This framework would allow for real-time navigation aid. This study proposes an architecture that efficiently captures local characteristics and global contextual information by integrating a reduced transformer encoder with a lightweight convolutional backbone. Structured pruning and low-precision inference are examples of model-level optimization that drastically cut down on memory usage and computational overhead, allowing for steady real-time performance. To improve domain relevance, training was carried out utilizing a two-stage approach that mixed a task-specific assistive navigation dataset with large-scale benchmark data (MS COCO dataset). Further enhancement in visual conditions, a multimodal sensing using ultrasonic distance estimation is needed to include. The final result of the work shows the optimized transformer framework with good comparison between the latency, detection accuracy, and resource consumption, which shows a suitable model for deployment on the continuous edge. Moreover, considering efficiency aware transformer models shows scalable assisted navigation systems which also boosts autonomous movement.

Keywords: Edge Computing, Transformer-Based Object Detection, Assistive Navigation, Real-Time Inference, Embedded Systems

I. INTRODUCTION

It is extremely dangerous for people with visual impairments to navigate independently, especially in both indoor and outdoor settings where there are constantly shifting obstacles, items in motion, and other forms of environmental unpredictability. Traditional forms of mobility assistance, including guiding dogs and white canes, have limited spatial awareness and cannot transmit detailed contextual information about the environment. Current improvements in computer vision shows camera based assistive navigation systems that detects obstacles and provides real time guidance to the system [2]. Nevertheless, combining this intelligence in portable, low-cost devices presence trade-off between the latency, computation, memory, energy consumption, refereed as edge constraints [3].

II. LITERATURE REVIEW

The authors, Nguyen et al. [6] introducing a lightweight transformer architecture with reduced parameters and optimized attention mechanisms for object detection on resource-limited edge surveillance systems. Further, authors Liu et al. in ref. [7] mentioned a low-complexity YOLO based object detector optimized for edge platform. This work enhanced the data augmentation, lightweight decoupled head, and a hybrid loss function shows 50.6% on MS COCO and real-time performance exceeding 230 FPS on NVIDIA Jetson AGX Xavier. Another article ref. [8], authors Okolo et al. developed a smart assistive navigation system for visually impaired users that integrates YOLO v8 based object detection with voice driven guidance, but this system lacks multimodal sensing. The authors Patel et al. in ref. [9] presented a complete comprehensive survey on navigation aids for visually impaired individuals, which emphasizing the role of multisensory integration. Most of the existing solutions prioritizes on accuracy of navigation or speed of the isolation, insufficient multimodel sensing, and system level sensing. So, we proposed a cohesive edge-based sensing and feedback pipeline for navigation system in real-time visual data.

ISBN: 978-81-985770-0-9

III. PROPOSED METHODOLOGY

System overview

The proposed navigation system functions as a cohesive edge-based sensing and feedback pipeline that operates in real-time. For objective detection and scene comprehension, forward-facing RGB camera functions as the principal sensing modality. The other sections show the edge-optimized transformer architecture, model optimization, muti-modal sensor fusion, audio feedback module, and its experimental setup

A. Edge optimized transformer architecture

The object detection framework employs an edge-optimized transformer architecture specifically designed to operate under strict computational and memory constraints. A lightweight convolutional backbone is utilized to efficiently extract low- and mid-level visual features, reducing redundant computations while preserving essential spatial information[10]. The specified features are then processed by a compact transformer module with a limited number of encoder layers, facilitating global context modeling with little computing expense. An adaptive query selection technique was integrated to boost performance by enabling the model to dynamically prioritize informative object queries according to scene complexity.

B. Model optimization

The detection framework incorporates various model-optimization algorithms to facilitate real-time execution on resource-limited hardware. Structured pruning is utilized to eliminate superfluous filters and attention components that have a negligible impact on detection performance, hence yielding a more compact and efficient network. Post-training quantization utilizing INT8 precision decreases computational complexity and memory use while preserving consistent inference accuracy.

C. Multi modal sensor fusion

The assistive navigation system employs a multimodal sensing approach to enhance dependability and environmental awareness. Vision-based object detection offers a semantic comprehension of the environment by recognizing obstacles and navigation-related entities from camera data. A late fusion method was utilized to integrate the outputs of both sensory modalities at the decision level, facilitating obstacle confirmation and minimizing false positives.

D. Audio feedback module

The audio feedback module converts perceptual outputs into intuitive auditory cues to facilitate safe navigation. The identified items are characterized by their type, estimated distance, and relative direction, allowing users to construct a precise mental image of their environment. Spatial cues are produced from integrated visual and ultrasonic data to guarantee precise and prompt reactions. A low-latency text-to-speech engine was utilized to provide succinct audio messages with negligible delay, maintaining real-time responsiveness during motion. The feedback technique emphasizes clarity and conciseness to avoid cognitive overload while ensuring situational awareness, rendering the system appropriate for ongoing use in dynamic and unpredictable settings.

E. Experimental setup

i. Hardware and training strategy

The experimental configuration was executed on low-power embedded systems often utilized for edge intelligence applications. A Raspberry Pi 4 and an NVIDIA Jetson Nano were utilized to assess computational efficiency, latency, and deployment viability across varying hardware capabilities, as illustrated in Table 1. Visual data was obtained using a conventional USB camera positioned to capture forward-facing scenes pertinent to navigation activities. An ultrasonic sensor was incorporated to deliver real-time distance readings for proximate obstructions, augmenting visual perception and improving system resilience under difficult settings.

ii. Training strategy

The model training employed a dual-phase learning strategy to guarantee both generalization and task specificity. Preliminary pretraining was performed on the MS COCO dataset to facilitate the acquisition of resilient object representations across varied situations. The pretrained model was subsequently fine-tuned with a task-specific assisted navigation dataset comprising

ISBN: 978-81-985770-0-9

indoor and outdoor navigation scenarios. The fine-tuning method customizes the detection model to specific item classes and spatial configurations typically found in assisted mobility settings, enhancing detection reliability in practical applications. The schematic representation of the proposed work is illustrated in Fig. 1. Table 2 presents the details of the dataset.

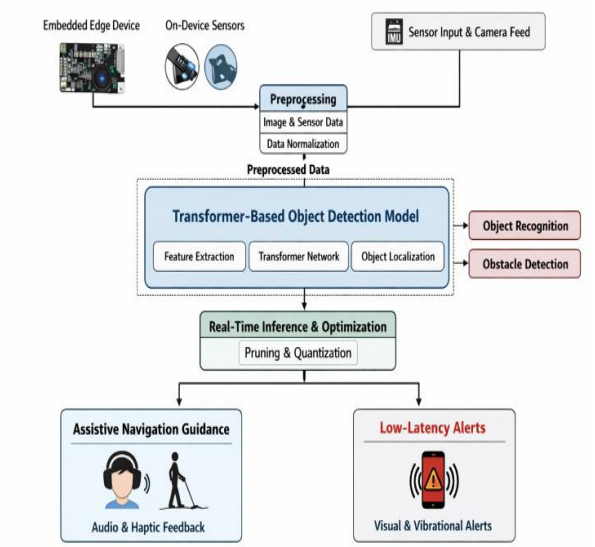

Fig. 1 Block Diagram of the Proposed Edge-Optimized Transformer-Based Assistive Navigation System

iii. Evaluation metrics

A performance evaluation was performed at both the detection and system levels to thoroughly test the efficacy of the proposed assistive navigation framework under real-time edge limitations.

TABLE 1.
HARDWARE COMPONENTS AND SPECIFICATIONS

Component	Description
Processing Unit	Raspberry Pi 4 / NVIDIA Jetson Nano
CPU / GPU	Quad-core ARM CPU / CUDA-enabled GPU
Camera	USB RGB Camera
Distance Sensor	Ultrasonic Sensor
Operating Mode	Real-time edge inference

TABLE 2
DATASET USED FOR TRAINING AND EVALUATION

Dataset	Purpose	No. of Images	Key Characteristics
MS COCO 2017 [11]	Pretraining	118,000	Diverse object categories, complex scenes
Assistive Navigation Dataset	Fine-tuning	2,000–5,000	Indoor and outdoor navigation environments

Detection accuracy was quantified using the mean Average Precision (mAP) as mentioned in Eq. 1.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} \quad AP_i \qquad (1)$$

where AP_idenotes the average precision for the i-th object class and Nrepresents the total number of classes. The real-time capability was evaluated using frames per second (FPS), which was computed using Eq. 2.

$$FPS = \frac{1}{T_{inf}} \qquad (2)$$

Where T_{inf} is the average inference time per frame. The system-level efficiency was assessed using latency, power consumption, and memory usage.

IV. RESULTS AND DISCUSSION

i. Detection performance

The detection performance was assessed to measure both generalization capabilities and task-specific efficacy across various datasets and model configurations. The training outcomes for both datasets are illustrated in Figures 2 and 3. Comparative experiments were performed utilizing the MS COCO dataset and a task-oriented assistive navigation dataset to evaluate the effects of domain adaptation. The detection accuracy was measured using the mean Average Precision (mAP) as defined in Equation 1. To contextualize the performance, the suggested edge-optimized transformer model was juxtaposed with lightweight state-of-the-art detectors, including YOLOv5-lite and Tiny-DETR. These baselines were chosen because to their extensive application in edge and embedded vision contexts.

The findings demonstrate that refining the assistive dataset markedly improves task-specific detection accuracy. In comparison to CNN-based and lightweight transformer baselines, the suggested architecture attains enhanced accuracy while preserving real-time inference

ISBN: 978-81-985770-0-9

speed, illustrating its appropriateness for edge-based assisted navigation applications.

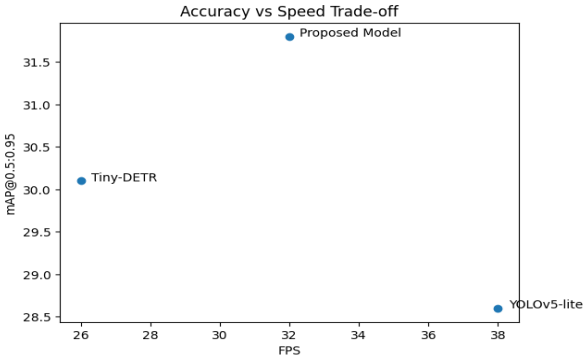

Fig. 2. Performance on MS COCO Dataset

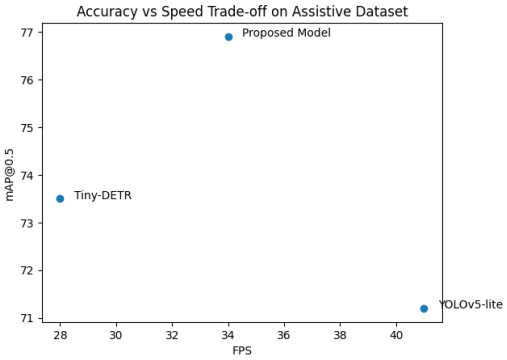

Fig. 3. Performance on Assistive Navigation Dataset

ii. Edge performance

Edge performance evaluation emphasizes measuring the efficacy of optimization strategies to enhance real-time execution and resource efficiency on embedded devices. The inference speed was evaluated prior to and after to the implementation of structured pruning and post-training quantization to determine their effect on frames per second (FPS), with the findings presented in Fig. 4. The modified model exhibited a significant enhancement in throughput, facilitating seamless real-time operation without sacrificing detection reliability. Alongside inference speed, power consumption and memory use were assessed to determine deployment viability on low-power edge devices. Power measurements were documented throughout continuous inference to ascertain the average energy consumption, whereas memory utilization indicated the runtime footprint of the model post-optimization.

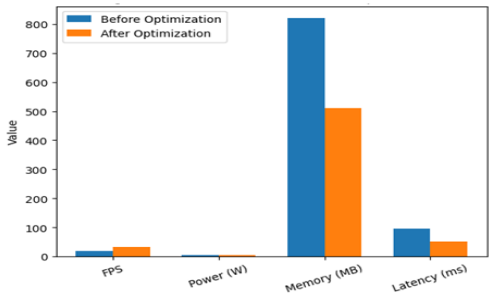

Fig. 4. Edge Performance Before and After Optimization

iii. Ablation study

An ablation study was conducted to analyze the individual impact of key architectural and optimization components on the detection accuracy and edge performance. The results of the ablation study are presented in Fig. 5.

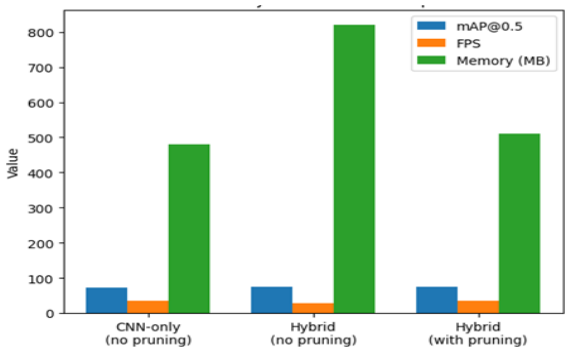

Fig. 5. Ablation Study Results

The initial investigation contrasted the model's performance with and without structured pruning to assess its impact on efficiency and prediction accuracy. The second research assessed the contribution of the hybrid convolutional-transformer architecture by contrasting it with a CNN-only baseline utilizing an equivalent lightweight backbone. The CNN-only model provides quicker initial inference but demonstrates diminished accuracy in intricate scenarios due to its constrained global context modeling.

V. CONCLUSION

This study demonstrates the viability of implementing transformer-based object detection models on resource-limited edge devices for assistive navigation purposes. The experimental findings indicate that the optimized

ISBN: 978-81-985770-0-9

transformer model achieves an average inference speed of 34 FPS and an end-to-end latency of roughly 52 ms on embedded hardware, validating its appropriateness for continuous navigation assistance. Memory utilization has been decreased to approximately 510 MB, while average power consumption stays under 4.5 W, facilitating continuous operation on economical, battery-operated devices. The system-level assessment further confirmed the real-time assistive navigation functionality by consistently identifying navigation-related objects in both indoor and outdoor settings. Fine-tuning on a task-specific assistive dataset enhanced detection performance by over 5% mAP over to generic pretraining alone, underscoring the significance of domain adaptation for effective deployment. The incorporation of multimodal sensing improves obstacle detection and diminishes failure occurrences in visually demanding situations.

REFERENCES

[1] A. Gudyś, M. Sikora, and Ł. Wróbel, "Separate and conquer heuristic allows robust mining of contrast sets in classification, regression, and survival data," *Expert Syst. Appl.*, vol. 248, 2024.

[2] C. Wang, K. Wu, and J. Liu, "Evolutionary Multitasking AUC Optimization [Research Frontier]," *IEEE Comput. Intell. Mag.*, vol. 17, no. 2, pp. 67–82, 2022,

[3] A. Ahmed *et al.*, "Multiple Power Line Outage Detection in Smart Grids: Probabilistic Bayesian Approach," *IEEE Access*, vol. 6, pp. 10650–10661, 2018.

[4] C. Sreekar, V. S. Sindhu, S. Bhuvaneshwaran, S. R. Bose, and V. S. Kumar, "Positioning the 5-DOF Robotic Arm using Single Stage Deep CNN Model," *Int conf on Bio Signals, Images, and Instrumentation*, 2021, pp. 1–6.

[5] F. Liu *et al.*, "Vision Transformer-based overlay processor for Edge Computing," *Appl. Soft Comput.*, vol. 156, p. 111421, 2024.

[6] D. Nguyen, V.-D. Hoang, and V.-T.-L. Le, "A Lightweight Transformer Model for Real-Time Object Detection on Edge Devices in Surveillance Camera Systems BT - Recent Challenges in Intelligent Information and Database Systems, 2025, pp. 221–234.

[7] S. Liu, J. Zha, J. Sun, Z. Li, and G. Wang, "EdgeYOLO: An Edge-Real-Time Object Detector," *Chinese Control Conf. CCC*, vol. 2023-July, pp. 7507–7512, 2023.

[8] G. I. Okolo, T. Althobaiti, and N. Ramzan, "Smart Assistive Navigation System for Visually Impaired People," *J. Disabil. Res.*, vol. 4, no. 1, pp. 1–10, 2025

[9] I. Patel, M. Kulkarni, and N. Mehendale, "Review of sensor-driven assistive device technologies for enhancing navigation for the visually impaired," *Multimed. Tools Appl.*, vol. 83, no. 17, pp. 52171–52195, 2024.

[10] Y. Xu, T. Xu, W. Zhang, J. Wu, Z. Cheng, and W. Yang, "Efficient Transformer-Based Visual Tracking for Edge Computing Devices BT - Artificial Intelligence and Robotics," 2025, pp. 223–232.

[11] T. Y. Lin et al., "Microsoft COCO: Common objects in context," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014,

Cite this article as:

Potipireddi Kumari and Mukku Pavan Kumar, "Real-Time Assistive Navigation Using an Edge-Optimized Transformer-Based Detection Framework", Proceedings of 13th international conference on Microelectronics, Circuits and Systems, Micro2026.
Displayed as online on 15th June 2026.

Link: http://actsoft.org/science/micro2026-pro/197-micro2026.pdf

ISBN: 978-81-985770-0-9