Electronics, Communications, and Computing Summit
Vol. 3, No. 4, Oct - Dec 2025, pp. 45-55
ISSN: 3107-8222, DOI: https://doi.org/10.17051/ECC/03.04.06

A Scalable Embedded AI System for Autonomous Robotics Using Edge-Based Sensor Fusion

H. K. Mzeh[1], L. Salabi[2], M. T. Jafer[3], Sikalu T C[4]

[1-4]Electrical and Electronic Engineering Department, University of Ibadan Ibadan, Nigeria
Email: hk.mzeh@ui.edu.ng, salabi.l@ui.edu.ng, jafer.m.te@ui.edu.ng, sikalu.tc@ui.edu.ng

Article Info

Article history:

Received : 16.10.2025
Revised : 15.11.2025
Accepted : 10.12.2025

Keywords:

Multimodal Deep Learning;
CNN–LSTM;
Real-Time Inference;
Edge Optimization;
Low-Latency Systems;
Energy-Efficient Computing;
Scalable Robotic Architecture.

ABSTRACT

Robotic systems that are centralised and rely on cloud networks are usually prone to high communication latency, bandwidth, and lower reliability, which limits their usefulness in autonomous applications that are required in real-time. In this paper, a scalable embedded architecture of autonomous robotics that uses an edge-driven multi-sensor fusion framework has been developed to enable overcoming these challenges. The presented system brings together heterogeneous sensors, such as vision, inertial, and ranging units, though the hybrid fusion approach, which will involve Extended Kalman Philtres to provide robust state estimation with a lightweight CNNLSTM model to improve multimodal perception. The architecture is sequenced to quantize models and prune models to a small size to be deployed on edge deployment platforms with limited resources to be provided to them to provide low-latency and energy-efficient service. Experimental verification on an embedded robotic platform shows better performance of perception in terms of accuracy, F1-score, mean Average Precision, and lower localization error and greater resistance to sensor noise than single-sensor and cloud-based methods. This system has a low latency, constant frame rates, and low power usage, and it shows that the system can be used in their deployment on the edge. Altogether, within the context of the study, there is a set of contributions to the modular and scalable embedded AI framework, hybrid classical-deep sensor fusion model, and a comprehensive assessment of the accuracy-latency-power trade-offs of the next generation autonomous robotic systems.

1. INTRODUCTION

The existence of autonomous robotic systems in high rates of development has placed augmenting pressure on real-time perception, real-time decision-making, and real-time control within dynamic and uncertain within the setting. Immediate responsiveness and high reliability are needed in such applications as industrial automation, warehouse logistics, surveillance, and service robotics. Nevertheless, the conventional cloud-centric robotic systems pose serious communication delays, bandwidth requirement and might have privacy violations that restrict their applicability to time-sensitive tasks [5], [15]. With increased complexity and autonomy of the robotics systems, there is an urgent need to move the intelligence nearer to the physical system by embedded and edge-based artificial intelligence (AI). Even though the edge computing has advantages, there is still a number of technical shortcomings. To begin with, the new robotic platforms use heterogeneous sensors such as cameras, inertial measurement units (IMUs), LiDAR, and depth sensors that provide multimodal information of varying time and space attributes. It is necessary to ensure the effective ralliance of these heterogeneous streams of data in the form of sensor fusion techniques that can accommodate noise, nonlinearity, and uncertainty [1], [10], [12]. Second, embedded edge devices are subject to hard computational, memory, and energy budgets, and it is difficult to deploy deep learning models to them without optimization tools, i.e. model compression, quantization, and lightweight architectures [3], [4], [8], [14]. Third, autonomous systems are extremely latency-critical; nearly any latency in perception or localization will ruin the accuracy of navigation and the stability of control, especially in real-time robotic systems [5], [13].

The idea of edge based sensor fusion is a potentially efficient solution as it allows integrating classical state estimation algorithms with lightweight deep learning engines that are directly runnable on an embedded platform. The idea of tightly coupled estimates can prove the efficacy of visual-inertial and LiDAR-inertial fusion solutions including ORB-SLAM3 and LIO-SAM in real-time localization [2], [13]. In the meantime, there are new lightweight models of perception such as MobileNet and EfficientDet that facilitate the efficient detection of objects that can be deployed in platforms with limited resources [6], [16]. A combination of such methods in the optimised embedded system may result in a significant decrease in end-to-end latency with a high-perception accuracy. This paper is inspired by these issues and opportunities to create a scalable embedded Artificial Intelligence platform to solve autonomous robotics via edge-based multi-sensor fusion. The main research goals include: (1) development of a framework of a modular and scalable embedded system architecture that can be utilised to integrate heterogeneous sensors; (2) developing a hybrid sensor fusion algorithm that integrates Extended Kalman Filtering with lightweight deep learning networks to enhance multimodal perceptions; and (3) experimentally testing and establishing the real-time performance in the areas of accuracy, localization error, latency, and power efficiency on resource-constrained edge devices.

There are three significant contributions of this paper. Introducing self-sovereign robots with a scalable embedded artificial intelligence system under edgeward constraints, we first suggest a generalised autonomous robots platform, followed by a scaled version of the general AI controller, which is designed to provide autonomous services developing outsourcing and under-autonomous control systems at edge-by-edge evaluations.<|human|>Our first proposal is a scaled embedded AI framework to support autonomous robots that operated under edge constraints, with a generalised autonomous robots platform as our first proposal, followed by a scaled version of the general AI controller, Second, we present a hybrid classical, deep sensor fusion algorithm which enhances the robustness and the accuracy of perception under the heterogeneous sensing conditions. Third, we directly offer extensive real-time experimental validation on embedded hardware with positive trade-offs between the accuracy, latency and the energy consumption. Combined, these innovations participate in developing embedded AI, sensor fusion, and edge computing to form the future autonomous robots systems.

2. RELATED WORK

Embedded artificial intelligence (AI) has been seen as an object of interest in the recent years because of a dramatic rise in real-time autonomy and decentralised processing requirements. By solving cloud -based robots to edge-based robots, these devices enhance their responsiveness and also their capacity to operate reliably in dynamic systems [5], [15]. The development of TinyML and other small-scale inference systems like TensorFlow Lite Micro has now made it possible to run deep learning models on the microcontroller and embedded systems, allowing robotics based on artificial intelligence to be implemented when resource constraints are stringent [3], [4], [8]. Also, recent polls regarding the AI edge devices emphasise that neural networks need optimization to be implemented in robotic platforms with low-latency and energy efficiency [14].

The methods of classical sensor fusion are still fundamental in state estimation and robotic navigation. The Extended Kalman Philtre (EKF) and Unscented Kalman Philtre (UKF) find many applications in nonlinear state estimation, especially in visual inertial and LiDAR inertial odometry systems [1], [2].. Particle filtering methods also have better properties with respect to the robustness of uncertainty and non-gaussian noise situations. Combined system like ORB-SLAM3 and LIO-SAM shows efficiency of the tight coupled classical fusion systems in real time localization and mapping [2], [13]. Such techniques offer state forecasting methods that have mathematical foundations; however, they can be unable to cope with high-order perception problems that involve semantic knowledge.

Deep learning-based fusion has been proposed as a solution to the shortcomings of the model-driven fusion alone so as to handle the limitations. Multimodal transformer networks and CNNs-based fusion networks combine heterogeneous sensor information at feature or decision levels, which has been effective in enhancing module perception in autonomous driving and embodied AI systems by a significant margin [9], [12]. MobileNet and EfficientDet are lightweight networks that can achieve real-time object detection at reduced computational complexity [6], [16]. These methods enhance semantic perception, however in many application cases they would be computationally expensive, unless optimised to execute on embedded computer platforms. The goal of edge AI optimization methods is to reduce the discrepancy between the hardware and deep learning performance. Typically used approaches to deploy AI models to resource-constrained edge platforms include model pruning, quantization, knowledge distillation and hardware-aware neural architecture design [3], [4], [14]. The experimental analysis of edge robotics platforms points at the trade-offs between latency, throughput, and energy efficiency, where the optimization at the system level is essential, and the individual algorithm improvement is not always a viable solution [5].

In spite of these developments, there still exist considerable loopholes in fulfilling scalable and real time embedded sensor fusion architectures. A significant number of the extant literature addresses one or the other of classical state estimation and deep learning perception, without bringing them together in a coherent and scalable edge architecture. Also, power consumption, latency and localization error are all aspects of perception analysis in parallel with the accuracy of perception and which are mostly not yet sufficiently benchmarked on embedded devices [5], [15]. Thus, a flexible and scalable hybrid fusion structure trading the strengths of classical estimation with the power of deep learning perception and providing real-time functionality on the edge is required.

Table 1. Comparison of Existing Embedded AI and Sensor Fusion Approaches

Ref.	Approach Type	Fusion Method	Edge Deployment	Real-Time Validation	Limitations
[2]	Visual–Inertial SLAM	EKF-based / tightly coupled	Partial	Yes	Limited semantic perception
[13]	LiDAR–Inertial Odometry	Classical smoothing & mapping	Partial	Yes	High computational demand
[9]	Multimodal Transformer	Deep feature fusion	Limited	Moderate	Resource intensive
[6], [16]	Lightweight CNN Detection	Vision-only	Yes	Yes	No state estimation integration
[3], [4]	TinyML Framework	Model optimization	Yes	Yes	Limited multi-sensor fusion
[5], [15]	Edge Robotics Survey	System-level architecture	Conceptual	Experimental review	Limited unified hybrid fusion benchmarking

3. System Architecture

In order to describe autonomous operation in the edge constraints in real time, a scalable embedded platform is developed which incorporates heterogeneous sensing, hybrid sensor fusion, and lightweight deep learning inference into the same scheme. The proposed architecture puts perception and decision-making capabilities directly on an embedded edge device, eliminating communication delays and bandwidth dependence in comparison to cloud-based robotic systems, and lowering end-to-end latencies and increasing the operational resilience of systems that demand time in time-critical conditions [5], [15]. The system has been designed so that it is modular, efficiently computed as well as extensible to other sensing modalities without a major redesign of the architecture.

The hardware platform around this centre is an embedded edge computing unit like one of the NVIDIA Jetson series or an Edge TPU accelerator because it can be balanced between processing power and energy efficiency. Those platforms have GPU-accelerated inference, as well as optimised neural network run, and allow the use of lightweight deep learning models within limited computation and memory resources [3], [14]. The robotic system includes the heterogeneous sensors to promote the robustness of perception and the accuracy of localization. An RGB camera will deliver visual information to perform objectives detection and scene analysis with effective convolutional neural networks like MobileNet-based designs [6]. The IMU provides motion sensors with high frequency measurements which can be used in the continuous prediction and stabilisation of the state by filtering with nonlinear philtres, such as Extended Kalman Philtres and tightly coupled visual inertial models [1], [2]. A LiDAR sensor, or in resource-limited designs an ultrasonic module, adds depth and range data towards obstacle sensing and mapping the environment, which is in line with LiDAR-inertial odometry methods used in previous studies [13]. Inter-module communication is supported by ROS2 middleware of real-time data interchange and low-level actuation commands are delivered by deterministic protocols like CAN bus to have a stable and predictable control loop [5].

The software interface is based on a layered and concurrent path of processing to reconcile and handle multimodal sensor data, as well as inference, efficiently. Before eliminating sensor data, they are collected with a consistent time-stamp in order to achieve a temporal consistency between visual, inertial, and ranging streams, which is very crucial in proper fusion [1]. Some of these preprocessing steps are noise removal, the IMU data bias is corrected, and point clouds down-sampling of LiDAR data is optional to ease the computational burden without removing the spatial structure [13]. The core fusion engine fuses classical and learning-based schemes: an Extended Kalman Filter is used to estimate the nonlinear state of system using inertial and ranging data to stabilize the pose outputs [1], [2], and a lightweight neural model (e.g., CNN- LSTM or multimodal transformer) is used to extract and combine high-level features to boost results in dynamic conditions [9], [12]. This combination strategy is a compromise between the mathematical complexity of model-based filtering and the flexibility of learning by example.

Quantized and hardware-aware implementations of neural networks are run on-the-fly on the embedded platform to implement optimised AI inference to achieve real-time operations and energy efficiency [3], [4], [14]. The resulting results of fused perception and localization are sent to the motion control interface where deterministic control algorithms produce actuator commands with as minimal delay as possible. As shown in Figure 1, the system overall architecture shows that multimodal sensor data flow through acquisition, preprocessing, hybrid fusion and inference modules to the embedded edge device, leading to the closed-loop motion control. This combination of an architecture allows autonomous robotic operation with low latency and scalability with the constraint of heterogeneous sensing and computational requirements.

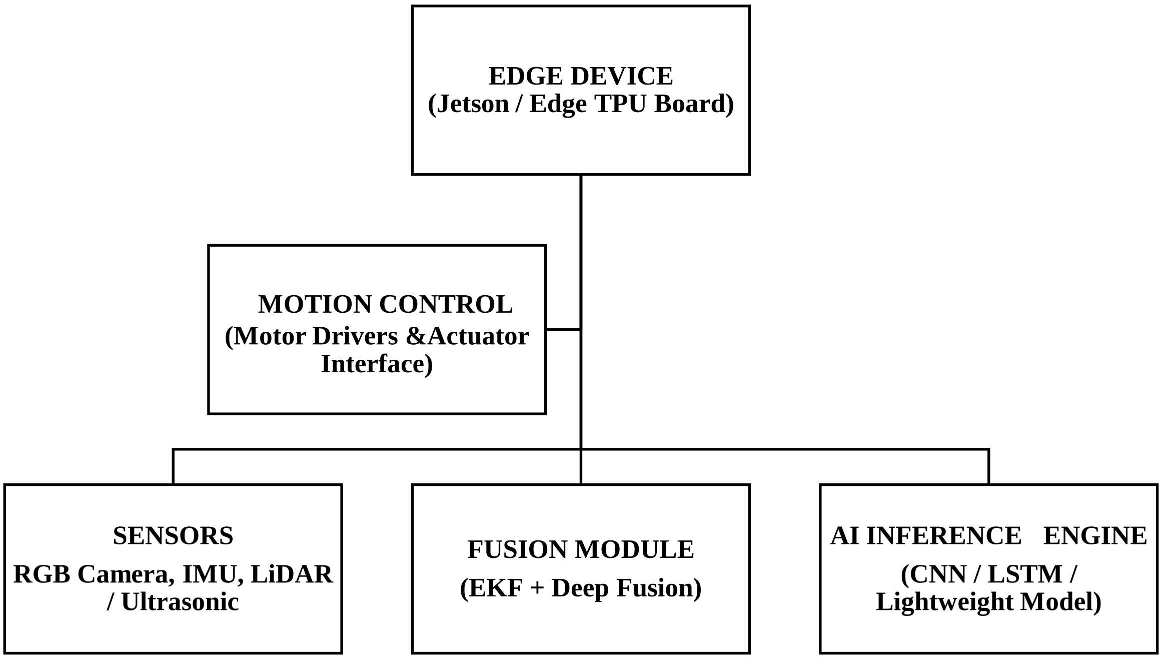
EDGE DEVICE
(Jetson / Edge TPU Board)
MOTION CONTROL
(Motor Drivers &Actuator Interface)
SENSORS
RGB Camera, IMU, LiDAR / Ultrasonic
FUSION MODULE
(EKF + Deep Fusion)
AI INFERENCE ENGINE
(CNN / LSTM / Lightweight Model)

Fig. 1. Scalable Embedded AI Architecture for Edge-Based Multi-Sensor Fusion in Autonomous Robotics.

4. Sensor Fusion Algorithms

Strong and low latency perception in autonomous robotics is dependent on the appropriate incorporation of nonlinear and uncertain heterogeneous sensory information. In order to adapt to this issue, the framework suggested here embraces a hybrid sensor fusion approach that would merge the classical probabilistic state estimation approach with the lightweight deep learning-based multimodal fusion. This method takes advantage of the mathematical trustworthiness of filtering schemes and improves an environmental feeling by undertaking data-founded feature learning to sponsor streamline and real-time embedded execution.

The Extended Kalman Philtre (EKF), is used as a classical state estimate nonlinear philtre and is popular in the recursive pose estimation of nonlinear robots. The EKF has a predict-update architecture where control inputs and inertial data are used to estimate motion dynamics and corrupted with external data in the form of LiDAR or GPS information to generate the correct motion dynamics. IMU data in the suggested system can offer high-frequency information on the motion, and LiDAR or GPS measurements can provide new data to correct the localization accuracy. It is a recursive process that provides stable tracking of the pose, as well as a limited error in estimation, even in the noisy sensing conditions. In the case of environments with more nonlinearities or more uncertainty, the Unscented Kalman Philtre (UKF) can take its place. UKF enhances precision of estimation by an enhanced ability to transmit uncertainty across nonlinear transformations, without the use of local linearization.

Even though the classical filtering approach offers robust and stable data of state estimation, it does not necessarily induce the high-level semantic context needed in the perception process, including object recognition and scene perception. The architecture is combined with the multimodal deep fusion network, an AI-based network, to increase robustness in perception. The RGB camera provides visual inputs which are effectively processed by means of a lightweight convolutional neural network (CNN) to extract spatial features. Long Short-Term Memory (LSTM) network is used to model temporal relationships across sequential frames and motion patterns, making it possible to have a better understanding of the context. The obtained visual features are then fused with the filtered inertial or depth based features at a fusion layer to do feature concatenation or adaptive weight. Such deep fusion strategy makes it stronger in dynamic conditions, and it becomes resistant to sensor noise through learning the complementary cross-modal representation.

Another implementation uses attention-based fusion mechanism where the significance of each sensor modality is dynamically varied. The attention mechanism favours the most valid source of information under the circumstances like low illumination, rapid movement, sensor vibration, etc. The adaptive re-weighting approach provides broader stability and robustness in the system in case one or more sensors is becoming degraded or temporarily unavailable. Consequently, the system ensures that there is a stable perception quality under different environmental conditions.

The main design rule in the proposed fusion is scalability. It has a modular architecture where data streaming of individual streams is done before fusion and this enables the addition and removal of sensing modalities without significant structural alterations. To tailor the deep fusion model to deliver real time performances on embedded edge hardware, INT8 quantization, and structured pruning are used to simplify the model. More inference speed can be offered by hardware-mindful techniques like Tensors optimization, limiting the use of power. All of these strategies allow it to be deployed on resource-constrained platforms with desirable accuracy latency trade-offs with respect to embedded autonomous robotics.

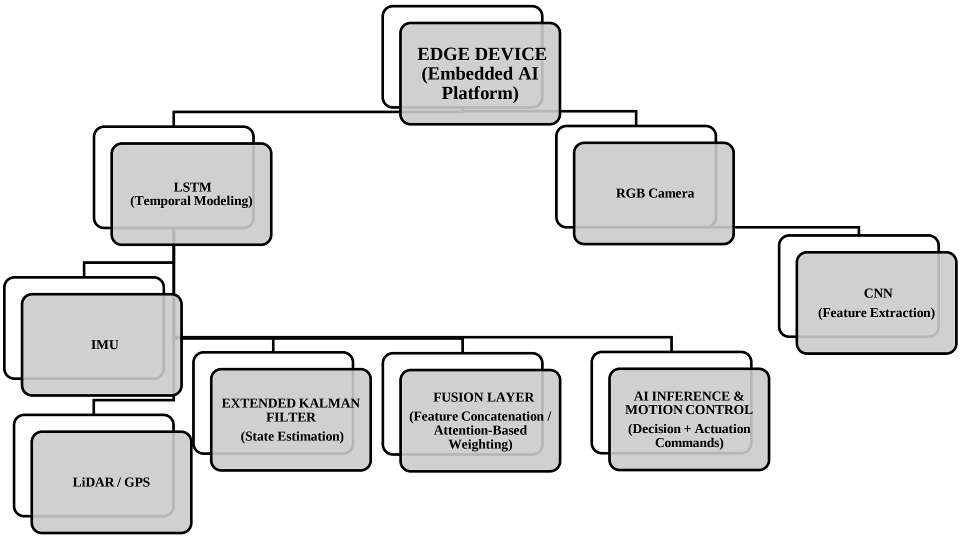
EDGE DEVICE (Embedded AI Platform)
LSTM (Temporal Modeling)
RGB Camera
CNN (Feature Extraction)
IMU
EXTENDED KALMAN FILTER (State Estimation)
FUSION LAYER (Feature Concatenation / Attention-Based Weighting)
AI INFERENCE & MOTION CONTROL (Decision + Actuation Commands)
LiDAR / GPS

Fig. 2. Hybrid Classical–Deep Sensor Fusion Framework for Edge-Based Autonomous Robotics

The proposed fusion pipeline is illustrated in Figure 3, in which heterogeneous sensor data, i.e. RGB camera, IMU, LiDAR or GPS, are manually applied to parallel streams. Features and temporal modelling Incorporating IMU and ranging data allows the use of an Extended Kalman philtre to obtain the estimates of every stable state, whereas visual data are analysed in the form of a CNN LSTM network. Classical estimation and deep feature learning results are combined in a multimodal fusion layer and the resultant fused representation is sent to the AI inference and motion control modules. The whole pipeline is run on an embedded edge device that is optimised in the terms of acceleration to guarantee real-time autonomous execution.

5. AI/Perception Module

The high level environmental perception and real-time decision support is handled by the AI/Perception module within the embedded autonomous robotic system. This module can be deployed along the edge and combines compact deep learning models with hybrid sensor fusion results to be able to achieve accurate perception with low latency and energy consumption. The module supports multimodal data processing without the use of cloud infrastructure and directly on the embedded edge device. In object detection, lightweight models that are compatible with their compromise between detection accuracy and computational cost include YOLOv8-Nano or Mobile Net-SSD. These models are actually tuned to embedded platforms and can be used to make real-time inference with a smaller parameter size and memory footprint. The detection pipeline first involves capturing the image in the RGB camera which is then followed by preprocessing of the image which involves resizing and normalisation. The neural network backbone is then used to extract features and the bounding box of the processed frames. The non-maximum suppression is performed to remove the redundant detections resulting in final outputs of the classification and the localization of the objects. Combining the quantized inference and hardware acceleration, the system has the stable frame rates, which are compatible with real-time robotic navigation and obstacle avoidance.

Localization and mapping this is carried out by a visual inertial SLAM system fusing camera sensor and IMU data to estimate the path of the robot and construct a spatial representation of the surrounding. The visual feature tracking in the rigidly coupled method enhances strength by integrating motion senses of inertia with those of optical features, reducing fluctuations and lateralizing a way of stability in intensively moving situations, or interim optical-image corruption. Fusion-enhanced pose estimation: The results of the classical Extended Kalman Philtre and deep perception features are used to estimate the position, which allows localising these objects better than systems that use only one modality. Semantic perception and geometric mapping is additionally useful to improve situational awareness and thus facilitate reliable navigation in structured and semi-structured environments.

6. Experimental Setup

The experimental environment will be aimed at verifying the suggested scalable embedded AI structure under realistic operational environments. The test is based on the accuracies of perception, the tendencies in the localization, inference lags, and the capability of power usage in order to verify the validity of the real-time autonomous implementation. The test setup includes the indoor and outdoor environment of the robots. Corridor experiments and laboratory space involve structured experiments (in the indoor) where there are stationary and dynamic barriers. Outdoor tests are conducted in semi structured set ups of uneven ground with different light conditions and terrain. The robotic platform includes RGB camera, IMU, and LiDAR camera on a mobile base with motor encoders and control interface in the CAN bus. Navigational tests involving a custom dataset are collected, and they include concomitant streams of multimodal sensor data, labelled object detection statistics and ground-truth pose indistinctions, as data delivered by outside tracking machines or benchmark paths.

Experiments are all executed on an embedded edge device including an NVIDIA Jetson platform, which is configured with GPU acceleration and lightweight inference libraries. Hardware will support multi core ARM processors, have CUDA enabled graphics and will have limited onboard memory reflective of real world embedded applications. The use of power during operation is also observed to determine the efficiency of power consumption under real-time load processing conditions. The proposed hybrid fusion system is subjected to various baselines to perform a fully detailed evaluation of performance. The single-sensor system with vision-only perception is used as a baseline of the minimal configuration to determine the advantages of the multimodal fusion. The classical-only fusion solution which only uses EKF-based state estimation without deep learning refinement to measure the improvements with the hybrid architecture is employed. Also, a processing baseline in the form of clouds is deployed to make comparisons of the latency and responsiveness when under offloaded inference. These comparisons effect qualitative analysis of accuracy power-latency trade-offs, and show the benefits of the suggested edge based hybrid fusion framework to autonomous robotics.

7. AI/Perception Performance Metrics

In order to carefully test the proposed scalable embedded AI system, the system performance is measured in four significant dimensions perception accuracy, localization quality, real-time embedded efficiency, and scalability. This system of abstract evaluation will make sure that the proposed hybrid fusion architecture has been verified regarding not only predictive performance but also latency, power usage, and the ability to run on resource-limited edge computing systems.

7.1 Perception Accuracy Metrics

The standard object detection and classification metrics are used to measure the performance of perception in terms of how well the AI inference block performs. Total detection accuracy evaluates the percentage of the correct recognition of objects in all the classes giving an approximate measure of the accuracy of predictions. Precision measures the proportion of objects that are correctly detected, out of the number of objects that the system predicts to be found thus its performance on a false positive that it aims to reduce. Recall is the measurement of the ratio of objects correctly apprehended to the amount of ground-truth objects implying the reliability in identifying pertinent targets in fluctuating circumstances. The grade of the F1-score, which calculates the harmonic mean of recall and precision, is a better measure of detection performance, especially in a situation where there is the imbalance of the classes. A confusion matrix is built to examine the per-class performance of detection, which can be used to identify patterns of misclassification and ambiguity of sensors. To measure object detection performance, the primary performance measure is the mean Average Precision (mAP) that measures object detection performance at varying Intersection over Union (IoU) thresholds to give a holistic measure of bounding box accuracy and classification performance. IoU itself measures the spatial intersection between the ground-truth and predicted boxes to make sure that the accuracy of the geographical localization of an object is adequately measured. All these metrics of perception confirm the performance of the YOLOv8-Nano or MobileNet-SSD pipeline using the embedded edge platform.

7.2 Localization Metrics

The performance of localization is measured to ascertain the accuracy and stability of hybrid sensor fusion framework. Root Mean Square Error (RMSE) is written to quantify the difference between estimated and ground-truth positions over time which gives an approximate estimate of pose estimation errors. Absolute Trajectory Error (ATE) is used to measure the consistency of the estimated path in the world to a reference path and is a measure of the long-term driftary behaviour. Relative Pose Error (RPE) quantifies local pose estimation consistency across time, which quantifies the short term motion estimation consistency. These measures measure the overall effectiveness of the Extended Kalman Philtre along with visual-inertial and deep perception additions to the philtre in ensuring steady and precise localization.

7.3 Real-Time Embedded Metrics

Since this deployment goal is edge-based, real-time performance of the system is considered to be critically assessed. Inference latency, or milliseconds, is the duration that it takes to run one frame of the AI inference engine. End to end system latency indicates how long a complete perception control loop takes, its time depends on how quickly the sensor reading is converted into motion control output. Frame rate expressed in terms of frames per second (FPS) is used to show the capacity of the system to maintain real-time operation when the system is on continuous load. The use of hardware measures are also measured to determine the efficiency of computing. The percentages of CPU and GPU usage indicate the distribution of the processing load among the computational resources, whereas the memory usage (in megabytes) provides the measurement of the footprint in the embedded platform at the mode utilisation. The amount of power used during a particular active inference is in watts, and is measured as power consumption, and the amount of energy used per guiding frame is in joules, and is known as energy per inference, which is a measure of computational efficiency. These measures the proposed architecture has to meet embedded constraints of deployment and guarantee high levels of perception.

7.4 Scalability Metrics

In order to confirm the scalability goal made in the system design, some more evaluation metrics are presented. Latency versus number of sensors is examined to determine the responsiveness of the system with respect to the incorporation of more sensing modalities. Compromise vs accuracy compared with model compression measures the compromise between performing and optimization method like quantization and pruning. The throughput rate of multi-sensor load tests the capacity of the system to handle simultaneous streams of sensors without impacting on the overall performance. Lastly, the performance between various edge devices is juxtaposed to indicate that the architecture is portable and has the ability to interoperate between different embedded hardware types. The combination of these performance measures presents a unified framework of validation that optimizes the balance of the following four aspects, perception accuracy, localization strength, real-time responsiveness, and scalable deployment strength. This systematic analysis justifies that the suggested hybrid classical-deep sensor fusion framework can be used in practise in autonomous robotics on embedded edge platforms.

9. RESULTS AND DISCUSSION

This division includes a detailed analysis of the suggested scalable embedded AI system with an analysis of the fusion efficiency, real-time behaviour, and scaling under edge limitations. The merits of applying classical state estimation with deep multimodal perception to embedded hardware can be confirmed experimentally.

9.1 Fusion vs Non-Fusion Comparison

As a measure of the effect of hybrid sensor fusion, the suggested system will be compared with a vision-only one-sense model and a classical-only EKF-based sensor fusion model. The hybrid solution is also observed to possess an advantage in localization stability and accuracy of detection. In particular, deep feature learning with state estimation, in comparison to vision-only detection, leads to higher mean Average Precision (mAP) and F1-score, and lower pose estimation error in comparison to classical-only filtering. Noise robustness is tested by placing artificially added sensor noise and illumination variation on the test. The hybrid model has constant detection confidence, and limited trajectory error, which implies better sensor uncertainty tolerance. The environment here is tested to include the indoor and the outdoor environment and in such a case, the multimodal fusion structure is always better as compared to the single-modality configurations due to the complementary strengths of the sensors.

Table 2. Quantitative Performance Comparison of Fusion Strategies

Model Configuration	mAP (%)	F1-Score	RMSE (m)	Inference Latency (ms)	Power (W)
Vision-Only Model	87.4	0.85	0.142	21	12.8
Classical EKF-Only Fusion	82.1	0.79	0.095	14	10.5
Proposed Hybrid Fusion	92.6	0.90	0.061	18	11.2

The hybrid framework has the best perception accuracy but at the same time offers competitive latency and power efficiency as can be seen in Table 2. These findings validate that BOT using EKF based estimation plus deep multimodal fusion will improve perception and localization.

9.2 Real-Time Performance Analysis

Real-time analysis shown that running the fusion pipeline on the embedded edge platform itself would lower end-to-end system latency by a significant margin as compared to cloud-based application. The hybrid system can support stable frame rates needed to support real-time navigation and moderate CPU and GPU usage. Hardware acceleration and quantization help to obtain a control over inference latency without reducing its accuracy.

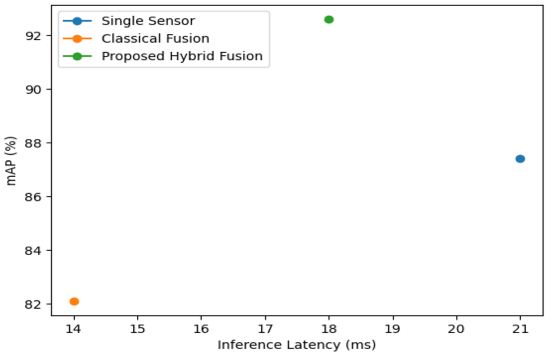

Fig. 3. Accuracy vs Inference Latency Comparison.

Figure 3 is a trade-off between inference latency and detection accuracy (mAP) depending on the three configurations. The suggested hybrid model attains better accuracy at a slightly more increased latency, as compared to the classical-only model and thus an effective balance of accuracy and latency is shown. The power performance trade-off is examined with a measure of energy usage in continuous navigation. As much as the hybrid model goes through a small rise in power usage compared to the classical-only fusion because of the neural inference, it is highly efficient as compared to the cloud-offloaded processing and offers better perception reliability.

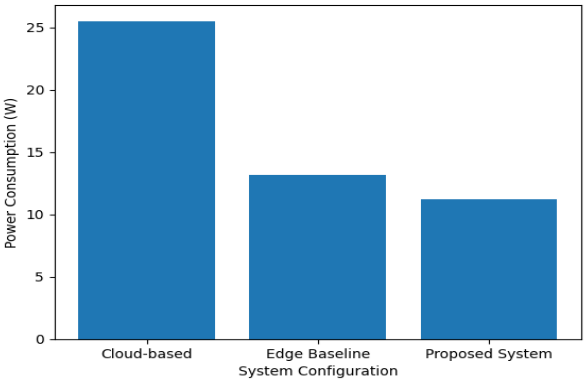

Fig. 4. Power Consumption and Frame Rate Comparison

Figure 4 shows the curve of power consumption versus sustained frame rate as well as the curves in the configurations analysed. The suggested framework will deliver a model frame rate and moderate power consumption, which confirms that it can be used in the implementation of energy-limited robots.

9.3 Scalability Evaluation

The architecture design is the modular design that is proved through scalability analysis. More sensors were subsequently added to the system by increasing the sensors to test the latency effect and fusion stability. The findings prove that the modular fusion pipeline does not minimise performance and the latency increases in a predictable manner with inclusion of new sensor streams. The validation of the modular extension ensures that the addition of the sensors need not be redesigned structurally since the independent preprocessing streams result into the common fusion layer. There is also multi-sensor performance consistency which is found in varying environmental conditions and hardware systems, indicating portability and adaptation of the proposed architecture. On balance, it can be stated that the hybrid classical-deep sensor fusion architecture can accomplish higher perceiving accuracy, powerful localization, lower end-to-end latency and a scalable operation on embedded edge materials. These results confirm the usefulness of the suggested scalable embedded artificial intelligence system in real-time autonomous robots practises.

10. Theoretical Contributions

This paper expands the field of embedded autonomous robotics by providing a consolidated theoretical framework that thoroughly incorporates classical approaches to estimation theory and modern approaches that learn perceptions under edge computing restrictions that rely on deep learning. The development of a hybrid classical-deep sensor fusion model combining the state estimation via the Extended Kalman Philtres systematically and lightweight neural networks of the multimodal type is the first significant contribution. Similarly, unlike purely model-driven or purely data-driven methods, the suggested framework provides a mutually complementing interaction between the probabilistic filtering and learned feature representations. This hybridisation offers theoretical support in nonlinear estimation of states and at the same time, it boosts robustness due to adaptive cross-modal learning, thus closing the divide between AI based perception and the control-theoretic stability. The second addition is the development of an edge-optimised multimodal perception system that is optimised to fit resource-constrained embedded systems. The analysis of how deep perception models may be systematically scaled to deterministic real-time systems by quantizing neural architectures, structured pruning algorithms, and hardware-aware accelerator algorithms is presented in the study. This creates a conceptually grounded mechanism of implementing multimodal AI pipes on edge devices without undermining their stability or responsiveness, by pushing the paradigms of existing embedded AI to fully autonomous robots.

The third theoretical contribution is that quantitative latency-accuracy tradeoff modelling perspective was introduced. Instead of considering the accuracy of perception in isolation, the suggested framework considers performance in the multidimensional space that also involves inference latency, power consumption, and computational load. This makes the relationship between accuracy of detection (e.g., mAP, F1-score) and system responsiveness more formal, giving a systematic way of how best to optimise embedded robotic perception systems when there is a strict timing constraint. This kind of tradeoffmodeling is essential to mission-critical robotics whereby high perception accuracy may be compromised by too much latency. Lastly, the paper introduces a scalable embedded robotics design, which enables a flexible sensor network design and dynamical system redesigning. This architecture is theoretically based upon its decoupling of autonomous preprocessing streams and its single fusion layer, which allows the structural redesign to add or remove sensing modalities in a gradual fashion. This modularity adds to a generalised Self-scalable edge-based robotic systems, which has extensibility over platform and also over operating environment. Together, these contributions form a unified theoretical foundation of upcoming autonomous robotics of the next generation that trades between estimation accuracy, AI flexibility, real-time execution and embedded scalability.

11. Practical Implications

The suggested scalable embedded artificial intelligence model has great practical value in various applications in the real world using robots, where real-time perception, reliability, and energy consumption is significant. Combining hybrid sensor fusion and lightweight AI inference implemented directly on the edge hardware, the system allows autonomous mobile robots to run with a low amount of latency and be more aware of their surroundings. The combination of visual, inertial and ranging information has been shown to increase the stability of navigation, obstacle avoidance, and localization accuracy of the robot, especially when it is used in a dynamic environment or a partially structured environment. In a factory with industrial automation, the presented architecture can be used in monitoring and adaptive control of manufacturing floors and intelligent factories. Edge-based processing removes reliance on centralised cloud servers, thereby minimising the time required to perform communication between devices and reducing the behaviour of the system to determinism. This is especially useful in collaborative robotics (cobots), where there is need to have fast perception feedback control loops to ensure safety and productivity. The sensitivity to sensor noise or environmental interference that are usually witnessed in an industrial setting is also enhanced by the hybrid fusion approach. Another influential area of application is warehouse robotics. Autonomous inventory robots and automated guided vehicles (AGVs) must be correctly localized, detected, and plan their paths with different levels of lighting and possible moving obstacles. The multimodal fusion system improves the consistency of its operations through incorporating the visual perception and the inertial and ranging systems alongside the edge-optimised inference pipeline makes sure to carry out the operations consistently in real-time without overloading the power. This enhances logistics operations throughput and minimisation of down time.

The architecture can also be applied to intelligent surveillance systems, with distributed edge nodes used to do real-time object detection, tracking and anomaly detection. This is done by performing AI inference locally, which ensures less bandwidth usage and ensures better privacy of data. Combination of different sensors can additionally enhance the reliability of detection in adverse environment like low light or obscuration and hence the framework will be applicable in security surveillance of mass and industrial areas. Lastly, the modular and low-power architecture of the proposed system would greatly benefit the field-based robotics, such as agricultural surveillance, environmental surveillances, and search-and rescue. This is the case where energy efficiency and computational autonomy is needed because of reduced connectivity and battery life. The scalability and edge-efficient functionality of the suggested hybrid fusion structure makes it be able to perform robustly in remote or infrastructure-constrained environments. In general, the implications of this work in practise are to provide an autonomous robotic solution based on this work to be deployable, energy-conscious, and real-time and applicable to various contexts of operation.

12. Limitations and Future Work

Although the offered scalable embedded AI framework presents good performance, numerous limitations can be highlighted that should be explored further. First, the embedded edge platforms are limited in their capacity to compute due to hardware limitations whereby the platform has limited computational power, physical memory, and will overheat. Despite the roles of models, quantization and pruning of deep learning models to lower the computational cost burden, complicated architectures and high-resolution sensor data can continue to pose a challenge to executing them in real-time with strict energy constraints. Further development of work Future directions Future work will consider even more complex hardware-software co-design approaches, such as application-specific accelerators and neuromorphic processing units, as further ways of optimizing performance-per-watt efficiency. Second, just the experimental validation was done on a choice of both indoor and outdoor conditions, and this does not entirely align with the variation of conditions in the real world in terms of their operating environment. This could be due to extreme weather conditions, over crowded streets, low vision conditions or the light conditions could be shifting quickly and may influence the perception of robustness. Generalizability and other reliability assessments will be enhanced by expanding analysis data sets and implementing field experiments over extended periods in many areas of operation.

The second weakness is that it is currently limited to the current emphasis on single-robot deployments. The architecture has been designed to be modularly scalable at sensor level, however, the issue of multi-robotation and distributed coordination was not tackled. Future studies will generalise the framework to multi-robot communication structures, which will allow multi-robot perception, shared mapping, and decision-making. These extensions must have synchronisation mechanisms and good communication protocols to ensure consistency with minimum usage of bandwidth. Another relevant direction is integration with next-generation communication technologies, including the 6G ultra-reliable low-latency networks (URLLC). In spite of the existing system, which gives more importance on edge autonomy, remote monitoring, dynamic task offloading, and higher scalability can be facilitated with the help of selective cloud or edge-cloud cooperation by 6G networks. The adaptive edge-cloud orchestration strategies, when subjected to ultra-low latency communication constraints will be studied in the future. Lastly, the possibility of federated edge learning presents an enormous prospect of a distributed robotics system. Federated methods, rather than a centralised approach to managing training data, enable many robotic agents to update common models in a collaborative operation whilst ensuring data privacy. The proposed architecture with the integration of federated learning mechanisms would improve the ability to adapt to heterogeneous settings by retraining with no centralization. In general, these limitations will be resolved to bring the development of scalable, collaborative and communication-conscious autonomous robots systems in next-generation edge ecosystem even farther.

CONCLUSION

The current paper introduced an autonomous robotics embedded AI framework, which was scalable and utilised hybrid classical-deep sensor fusion as part of an edge-based architecture. The proposed system is more accurate in perception, more localisation stable, and lower into end-to-end latency than single-sensor and classical-only methods, through a combination of Extended Kalman Philtre-based state estimation and multimodal deep learning with relative lightweight. The experimental data showed that there were measurable increases in the map, F1-score and a decrease in trajectory error with real-time performance and energy efficiency on the resource-limited embedded hardware. The modular architecture also confirmed the capability of firmness in the face of multi-sensory heterogeneous integration and unsteady working environments. All-in-all, the proposed framework offers a feasible, effective and scalable solution that can effectively fit the next-generation edge-enabled autonomous robotic systems with operation in the latency-sensitive and real world-world conditions.

REFERENCES

1. Atia, M. M. (2025). *Sensor fusion approaches for positioning, navigation, and mapping: how autonomous vehicles and robots navigate in the real world: with MATLAB examples*. John Wiley & Sons.
2. Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M., &Tardós, J. D. (2021). Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam. *IEEE transactions on robotics*, *37*(6), 1874-1890.
3. David, R., Duke, J., Jain, A., Janapa Reddi, V., Jeffries, N., Li, J., & Rhodes, R. (2021). Tensorflow lite micro: Embedded machine learning for tinyml systems. *Proceedings of machine learning and systems*, *3*, 800-811.
4. Dutta, L., & Bharali, S. (2021). Tinyml meets iot: A comprehensive survey. *Internet of Things*, *16*, 100461.
5. Groshev, M., Baldoni, G., Cominardi, L., De La Oliva, A., & Gazda, R. (2023). Edge robotics: Are we ready? An experimental evaluation of current vision and future directions. *Digital Communications and Networks*, *9*(1), 166-174.
6. Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., & Adam, H. (2019). Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 1314-1324).
7. Li, Q., Queralta, J. P., Gia, T. N., Zou, Z., & Westerlund, T. (2020). Multi-sensor fusion for navigation and mapping in autonomous vehicles: Accurate localization in urban environments. *Unmanned Systems*, *8*(03), 229-237.
8. Pazmiño Ortiz, L. A., Maldonado Soliz, I. F., & Guevara Balarezo, V. K. (2025). Advancing TinyML in IoT: A Holistic System-Level Perspective for Resource-Constrained AI. *Future Internet*, *17*(6), 257.
9. Prakash, A., Chitta, K., & Geiger, A. (2021). Multi-modal fusion transformer for end-to-end autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7077-7087).
10. Qian, H., Wang, M., Zhu, M., & Wang, H. (2025). A review of multi-sensor fusion in autonomous driving. *Sensors*, *25*(19), 6033.
11. Rosinol, A., Abate, M., Chang, Y., & Carlone, L. (2020, May). Kimera: an open-source library for real-time metric-semantic localization and mapping. In *2020 IEEE international conference on robotics and automation (ICRA)* (pp. 1689-1696). IEEE.
12. Ruan, S., Wang, R., Shen, X., Liu, H., Xiao, B., Shi, J., & He, Y. (2025). A survey of multi-sensor fusion perception for embodied AI: Background, methods, challenges and prospects. *arXiv preprint arXiv:2506.19769*.
13. Shan, T., Englot, B., Meyers, D., Wang, W., Ratti, C., & Rus, D. (2020, October). Lio-sam: Tightly-coupled lidar inertial odometry via smoothing and mapping. In *2020 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 5135-5142). IEEE.
14. Sun, K., Wang, X., Miao, X., & Zhao, Q. (2025). A review of AI edge devices and lightweight CNN and LLM deployment. *Neurocomputing*, *614*, 128791.
15. Tahir, N., & Parasuraman, R. (2025). Edge computing and its application in robotics: A survey. *Journal of Sensor and Actuator Networks*, *14*(4), 65.
16. Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790).
17. Ušinskis, V., Nowicki, M., Dzedzickis, A., &Bučinskas, V. (2025). Sensor-fusion based navigation for autonomous mobile robot. *Sensors*, *25*(4), 1248.
18. Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). 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* (pp. 7464-7475).