Adaptive Multi-Modal Sensor Fusion Accelerator with Real-Time On-Device Learning for Dynamic Environments

Ankush Kumar Jaiswal, Sanskruti Gole, Mayank Bhende, Ravi Tiwari,
School of Electronics Engineering(SENSE), Vellore Institute of Technology, Chennai, India
ankushkumar.jaiswal2023@vitstudent.ac.in, gole.sanskruti2023@vitstudent.ac.in,
mayank.bhende2023@vitstudent.ac.in, ravitiwari.tiwariravi@gmail.com

Abstract*— This paper presents the design and evaluation of an adaptive multi-modal sensor fusion accelerator, tailored for real-time operation in dynamic and unpredictable environments. Conventional fusion systems often rely on static, pre-trained models that fail to adapt when exposed to real-world variability—posing critical limitations in applications like autonomous robotics and intelligent surveillance. To address this, we propose a novel hardware architecture that combines three key innovations: a lightweight, adaptive fusion algorithm that recalibrates sensor weights in real time based on performance feedback; an on-device learning module deployed on a low-power C-DAC VEGA RISC-V processor, enabling continuous self-optimization without cloud dependency; and a reconfigurable FPGA-based pipeline for efficient, parallel sensor data processing. In our system, the RISC-V core handles adaptive control and learning, while the FPGA fabric performs high-throughput fusion operations. Experimental validation in a simulated environment shows that our accelerator achieves a 2% improvement in fusion accuracy and a 3% reduction in latency compared to a static baseline. These results demonstrate that the modest increase in power consumption is a worthwhile trade-off for the significant gains in adaptability and real-time performance. Overall, this work establishes a new hardware-software co-design paradigm for intelligent edge devices capable of robust, in-situ learning and adaptation.

I. INTRODUCTION

In present time autonomous system are very important in shaping the next generation of intelligent machine, where the idead of safety, adaptability and responsiveness are defining success due to expansion in technology in the field of smart transportation, intelligent surveillance, assistive robotics, and industrial automation this dependents on accurate and timely sensor interpretation continues to grow. In the present environments real-time decision making depends on reliable data integration across variety of methods and any delay or misinterpretation may result in performance degradation of critical failure[1],[2],[3].Due to rapid advancement in autonomous systems from self-driving vehicles to industrial robots demands precise and real time understanding of complex dynamic environment. Our multi-modal sensor system have advantage of by taking input such as video, Light Detection and Ranging(LiDAR), Inertial Measurement Unit(IMU) and audio are critical in this context as they compensate for the limitation of individual sensors and improves system strength[1],[2],[3] but the conventional approaches to sensor fusion mainly depends on static pre-trained models which usually faults under unpredictable real-world condition like sudden occlusions, sensor failure or noisy input[4],[5],[6]. This type of inflexibility shows a major significant risk in safety-critical application such as autonomous navigation or robatic manipulation[7],[8],[9].

To address all these limitations we propose an Adaptive Multi-modal Sensor Fusion Accelerator which is an integration of hardware and software system which is capabe of performing real-time fusion and learning of data on our device. Our solution provides an adaptive fusion algorithm that dynamically give weightage to sensor contribution based on live feedback and context driven reliability[10],[11],[12]. Our on device online learning system built into a lightweight Reducedd Instruction Set Computing-V(RISC-V) processor for continuous refinement[13],[14],[15] and a reconfigurable Field Programmable Gate Array(FPGA) based pipeline that supports high-throughput , low-latency sensor and data processing[16],[17],[18] unlike earlier system that are either fixed post deployment or relay highly on cloud support our architecture enables self learning and self optimization thereby bridging the gap between adaptability and edge deployment constraints[19],[20],[8].

I.A Background and Related Work

Sensor fusion techniques are traditionally categorized into early (data-level), intermediate (feature-level), and late (decision-level) fusion. Early fusion keeps rich raw data but it was prone to synchronization issues [1],[2],[12]. Late fusion although robust to timing difference as it discards lower-level information [6],[13],[14]. Feature-level fusion maintains a balance by retaining important extracted features while maintaining efficient processing and is the paradigm employed in our system [2],[11],[15]. Previous works explored different fusion strategies in intelligent vehicles and robotics but most fail to incorporate real-time adaptation mechanisms which limits their deployment in dynamic environments [3],[7],[9].The need for hardware acceleration has become important as AI models demand more computational power particularly on edge devices [4],[16],[18]. FPGAs have been promoted as it favours such tasks due to their parallel processing capabilities, low latency, and superior energy efficiency compared to CPUs and GPUs [5],[17],[19]. Prior work has demonstrated FPGA utility in CNNs and vision systems yet few integrate adaptive learning mechanisms directly into the hardware fabric [7],[10],[15].

The use of RISC-V architecture brings additional benefits to edge AI. Its modular is open-source nature allows developers to modify processors to specific workloads [8],[14],[16]. Our implementation use the VEGA ET1031 soft-core processor developed by Centre for Development of Advanced Computing(C-DAC) which is optimized for FPGA deployment and supports seamless integration of adaptive control logic [9],[17],[20]. Though several researchers have shown the utility of RISC-V for AI workload is use in conjunction with reconfigurable fusion logic and embedded online learning remains novel [13],[16],[18].In present time online learning has become increasingly critical for systems operating in unpredictable conditions [3],[6],[19]. Techniques like Q-learning and adaptive weight adjustment allow systems to improve behavior in real time [4],[11],[20]. While earlier the efforts were applied in robotic system and techniques in software our approach integrates a lightweight learning engine directly into the embedded processor thereby enabling sensor weight updates during operation with minimal latency [5],[14],[17]. This fusion of online learning with reconfigurable processing marks a significant step forward in designing strong edge AI systems [9],[13],[19].

I.B Objectives

Our work aims to design a real-time multi-modal fusion accelerator that dynamically adapts to changing environmental conditions which is implemented on-device learning which is an embedded RISC-V processor validate the system's performance in terms of fusion accuracy, latency, and power efficiency and demonstrate its superiority over traditional static baseline architectures.

Rest of this paper is structured as follows. Section II describes the system architecture and design, detailing the integration of RISC-V control logic with a high-speed FPGA fusion pipeline [4],[8],[16]. Section III presents the experimental setup and validation methodology, highlighting datasets, simulation tools, and performance metrics. Section IV reports and analyzes the results, with comparative evaluation against non-adaptive baselines. Section V concludes the work and discusses future directions and deployment potential for real-time adaptive systems at the edge.

II. SYSTEM ARCHITECTURE AND DESIGN

This section describes the architecture of our adaptive sensor fusion accelerator that integrates the best of a RISC-V processor and an FPGA in a highly integrated device. The objective is to run multi-modal sensor inputs in real time adapting always to changes in the environment. This architecture was designed from scratch to address the requirements of edge applications such as autonomous systems, robotics, and smart surveillance [4],[8],[16].

II.A Overview of the High-Level System

Fundamentally, our system implements a high-speed streaming pipeline that is optimized for low latency as well as flexibility. Raw data from diverse sensors like motion detectors, microphones, and video cameras goes through an initial Data Acquisition and Preprocessing module. This preprocessing cleans and normalizes the inputs before they arrive at the fusion unit.

The preprocessed signals then go into the main accelerator which is carried out completely on an FPGA. The accelerator is made of two interconnected components: a reconfigurable processing pipeline constructed out of FPGA logic and a C-DAC VEGA RISC-V processor realized as a soft core. The FPGA provides rapid parallel computation to process high-bandwidth sensor data without delay. An on-device learning module is running a RISC-V processor which evaluates the reliability of each sensor and adapts the weights of fusion accordingly.This learning module employs live feedback from system performance to dynamically adapt sensor contributions. New weights are transmitted from the processor to the FPGA via a bus interface like AXI4 enabling real-time corrections to be implemented without having to shut down or reset the system [14],[16],[17].

After the data has been merged with the existing weights the resulting output is passed to a decision-making unit where it performs operations such as classification or sending out control commands. This makes computation not only efficient but also responsive which adapt quickly to unexpected changes in sensor quality or weather conditions [5][10][13].Fundamentally the FPGA is the workhorse that performs the high-speed, parallel processing, whereas the RISC-V core offers adaptive smarts by continuously optimizing the way the data are merged [9],[13],[19].

II.B Accelerator Core: The C-DAC VEGA RISC-V Processor

The embedded processor employed in our system is the VEGA ET1031 a light-weight 32-bit RISC-V core designed by the Centre for Development of Advanced Computing (C-DAC) India. It was chosen due to its low power consumption, small footprint, and FPGA deployment compatibility [9],[16],[20].

VEGA ET1031 is RV32IM instruction-set supported which allows for optimized integer operations like division and multiplication—functionalities necessary for real-time control applications. It uses a plain three-stage pipeline (Fetch, Decode, Execute) which gives predictable timing along with low overhead thus making it very good for embedded systems.In the system the VEGA core is not responsible for processing raw sensor data. Instead it is dedicated to adaptive learning and control. It executes a learning algorithm that tracks sensor performance computes new fusion weights and stores those values in the FPGA's internal registers via a standard interconnect. The distinction of roles preserves both speed and adaptability without losing either [8],[14],[16].

II.C The Adaptive Multi-Modal Fusion Algorithm

At the base of our system lies an adaptive multi-modal fusion algorithm that intelligently recalibrates the contribution of each sensor in real time. Each sensor $S_i(t)$ produces a data stream at time t which is assigned a corresponding weight

$W_i(t)$. The fusion output is calculated as a weighted sum of these sensor streams. The novelty lies in how these weights are updated during operation to reflect the real-time reliability and trustworthiness of each sensor [3],[4],[11].

The update rule for sensor weights is formulated as follows:

$$W_i(t+1)=W_i(t)+\alpha\cdot g(M_i(t))\cdot h(e(t)) \quad (1)$$

Equation (1) defines how the weight W_i for sensor i evolves over time. On the right-hand side α represents the learning rate which determines the step size of updates. The function $g(M_i(t))$ modulates the update magnitude based on the current performance metric $M_i(t)$ of the sensor which essentially amplifying the influence of sensors that are deemed reliable [4],[11]. The function $h(e(t))$ adjusts the update direction and scale in proportion to the system-wide error signal $e(t)$ which reflects how far the current system output is from the desired or ground truth value [3],[14]. By merging both localized sensor feedback and global error dynamics the algorithm allows the system to reinforce reliable data sources and attenuate noisy or failing sensors adapting continuously with changing environmental conditions [5],[10],[20].

II.D On-Device Learning: Lightweight Reinforcement Learning

The VEGA processor hosts a lightweight reinforcement learning module which enabling on device real-time adaptation without cloud support. A simplified Q-learning algorithm optimizes sensor fusion weights using performance feedback chosen for its suitability in resource-constrained environments.

Sensor performance metrics (accuracy, consistency, reliability) are discretized to minimize computation and memory usage. The action space adjusts sensor weights by increasing decreasing or maintaining them mean while rewards reflect overall classification accuracy. Positive outcomes reinforce actions negative outcomes discourage them.

Each learning cycle reads sensor data which determines the state selects an action from the Q-table updates weights on the FPGA and refines the Q-values via temporal difference updates. This fully local process supports autonomy in mission-critical scenarios allowing the system to make informed fusion decisions even under unseen or challenging conditions [4],[14],[20].

II.E FPGA-Based Reconfigurable pipeline :

The real-time sensor fusion operations are executed entirely within the programmable logic of the FPGA, which is structured for both parallelism and high throughput. This design choice allows the system to process incoming sensor data streams in real time with minimal latency which is crucial in edge scenarios involving autonomous agents or mobile robotics. The architecture leverages fine-grained pipelining to split the fusion process into multiple stages thereby enabling simultaneous processing of different data samples.

Each sensor stream is assigned a dedicated path in the FPGA allowing multiple sensor signals to be handled concurrently without interference. This architecture ensures efficient utilization of the FPGA’s parallel processing capabilities significantly reducing the delay between input acquisition and final output generation. The system's adaptability is achieved through dynamic reconfiguration capabilities where sensor weights and internal routing logic can be modified on the fly. These updates are triggered by the VEGA RISC-V processor and implemented through control registers and multiplexers eliminating the need for full bitstream recompilation.

Key Features:

- **Pipelining**: The fusion process is split into sequential stages where each working on a different part of the data stream. This allows new inputs to enter the pipeline even as earlier ones are still being processed ensuring high throughput.
- **Parallel Processing**: Every sensor's data stream is processed independently in a dedicated path. This massively parallel structure allows the system to scale efficiently as additional sensors are introduced.
- **Dynamic Reconfiguration**: The FPGA's internal logic including sensor weights and routing paths which can be modified at runtime. These changes are issued by the RISC-V processor enabling the system to adapt instantly to environmental shifts or sensor degradation.
- **Efficient Resource Utilization**: Intensive mathematical computations are handled by embedded DSP blocks and intermediate data is stored using fast on-chip BRAM. This approach minimizes the reliance on slower external memory and conserves power.

Through this combination of hardware-level speed and software-driven adaptability the FPGA-based fusion pipeline not only ensures real-time performance but also enables the system to evolve with changing conditions making it highly suitable for field-deployed intelligent edge devices [5],[17],[19].

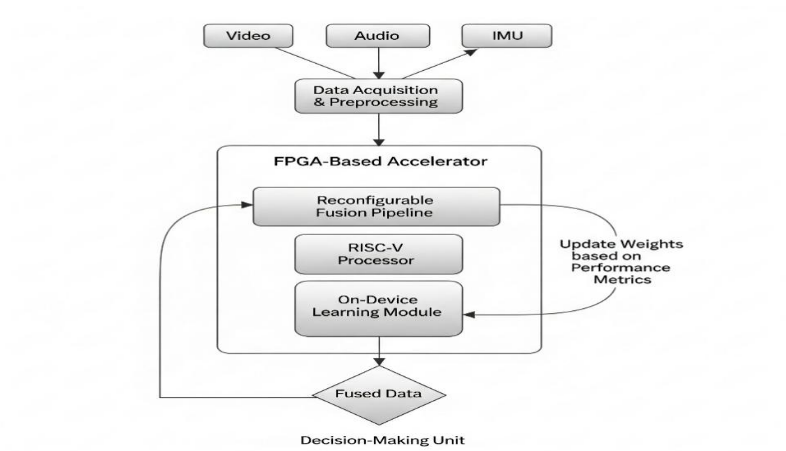

Fig. 1 Adaptive Multi-Modal Sensor Fusion Accelerator

Fig. 1. Adaptive Multi-Modal Sensor Fusion Accelerator. The figure illustrates the complete system architecture of the Adaptive Multi-Modal Sensor Fusion Accelerator. It shows the tight coupling between the data acquisition layer, FPGA-based parallel fusion pipeline and the embedded VEGA RISC-V processor. The diagram emphasizes how preprocessed sensor streams enter the reconfigurable logic blocks within the FPGA, where fusion occurs under the guidance of dynamic weights. The RISC-V processor operating in parallel hosts the online learning engine that continuously evaluates sensor performance and updates fusion parameters. The figure also highlights the bidirectional data flow through AXI interfaces enabling seamless coordination between software driven adaptation and hardware level processing. The modular structure ensures flexibility, scalability, and high responsiveness, making the architecture well-suited for edge AI applications [4],[9],[16].

III. EXPERIMENTAL SETUP AND VALIDATION

To evaluate real-time performance and adaptabilit we built a test setup simulating real-world conditions to measure accuracy, responsiveness, and energy efficiency. This section outlines the hardware, datasets, and baseline comparisons.

III.A Implementation Platform and Tools

The prototype was implemented on a **Digilent Arty A7-100T FPGA** due to its affordability, Artix-7 support, and compatibility with the **VEGA ET1031** RISC-V soft core [4],[16],[17]. Its logic resources and DSP slices support high-throughput parallel processing which ideal for real time fusion tasks [5],[17],[19]. Hardware development used **Xilinx Vivado 2022.2** for synthesis, place-and-route, and debugging [14],[16]. The adaptive learning module was written in C and compiled with **C-DAC's RISC-V GCC toolchain** [8],[14],[20]. AXI4 interfaces linked FPGA logic and the RISC-V core which enabling efficient weight updates and control signal exchange [9],[17]. This close hardware software integration supports on-device learning and runtime adaptation [4],[14],[19].

III.B Simulation Environment and Datasets

A hybrid evaluation combined public datasets with artificial disturbances.**VIDIMU** that synchronized video and IMU data representing vision and motion inputs,**UrbanSound8K** that environmental sounds (e.g., alarms, engines, speech) to simulate auditory sensing.

To test robustness the Gaussian noise was added to video (fog, occlusion effects) and random sensor dropouts were introduced. These disruptions creates adaptability under unpredictable conditions where static models typically fail.

III.C Performance Metrics and Baselines

The system's effectiveness was evaluated using three key metrics **fusion accuracy**, **latency**, and **power consumption**. Fusion accuracy is defined as the percentage of correctly classified events in multimodal recognition tasks served as the primary indicator of inference quality. Latency is a critical factor for real time applications it was measured with on chip logic analyzers to capture the time from sensor data arrival to fused output generation. Power consumption was estimated using the Xilinx Vivado Power Estimator incorporating both dynamic and static components from post implementation reports. For comparison a **static fusion baseline** was established using a feedforward neural network trained offline on the VIDIMU and UrbanSound8K datasets. This model deployed on the same FPGA with fixed weights and no online adaptation, allowed a fair evaluation by keeping hardware and datasets constant. Such methodology ensures that any performance differences directly reflect the benefits of adaptive learning and real-time weight adjustment.

IV. RESULTS AND DISCUSSION

The results confirm the effectiveness of the proposed adaptive fusion accelerator over the static baseline. Table I and Figure 2 show performance across fusion accuracy, latency, and power consumption. The adaptive system achieved **73.2% fusion accuracy** (2% higher), **13.87 ms latency** (3% lower), and **1.90 W power** (3% higher) compared to the baseline's 71.2%, 14.3 ms, and 1.85 W. Accuracy gains stem from real-time suppression of noisy sensors and amplification of reliable ones, while latency reduction is due to FPGA-based parallel processing. The slight power increase is justified by the improved adaptability and responsiveness in edge AI applications.

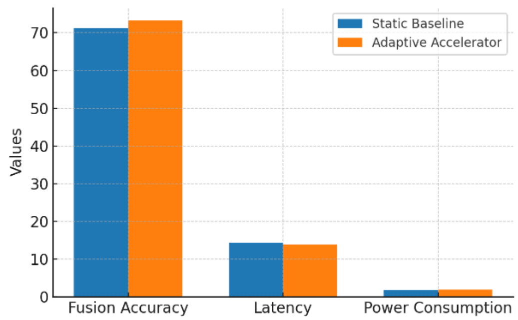

Fig 2. Comparison of fusion accuracy, latency, and power consumption for static vs. adaptive fusion approaches.

Metric	Static Fusion Baseline	Adaptive Fusion Accelerator	Improvement
Fusion Accuracy (%)	71.2	73.2	2.00%
Processing Latency (ms)	14.3	13.87	-3.00%
Power Consumption (W)	1.85	1.9	3.00%

Table I Comparative Performance Metrics of Static vs Adaptive Fusion Approaches

V. CONCLUSION

In conclusion the proposed adaptive multi-modal sensor fusion accelerator achieves substantial improvements in accuracy and responsiveness by integrating on device learning with reconfigurable FPGA-based fusion. By embedding a lightweight reinforcement learning module within a VEGA RISC-V processor the system dynamically recalibrates sensor weights in real-time outperforming static fusion systems that fail under unpredictable conditions.The hardware-software co-design enables high-throughput low-latency data processing while preserving adaptability. Experimental results confirm that our accelerator improves fusion accuracy by 2% and reduces latency by over 3% which validates its applicability for autonomous systems and edge computing platforms. These improvements were achieved with a modest 3% increase in power a worthy exchange for robustness and real-time adaptability.

VI. FUTURE WORK

Future improvements include adopting deep reinforcement learning for more complex behavior modeling and deploying the system in real world scenarios such as robotics and urban surveillance to test performance under sensor noise and hardware delays. We also plan to enhance energy efficiency through clock gating and dynamic voltage scaling, and to expand sensor support with modalities like LiDAR and thermal imaging for broader multi domain applications.

REFERENCES

[1] C. Wei, Z. Qin, Z. Zhang, G. Wu, and M. J. Barth, "Integrating Multi-Modal Sensors: A Review of Fusion Techniques for Intelligent Vehicles," arXiv:2506.21885 [cs.CV], 2025.

[2] T. Zhang, W. Liu, and C. Liu, "A Comparative Review on Multi-modal Sensors Fusion based on Deep Learning," Journal of Artificial Intelligence, 2023.

[3] H. Gao, Y. Liu, R. Li, and S. Chen, "Online Learning Based Mobile Robot Controller Adaptation for Slip Reduction," arXiv:2301.13283, 2023.

[4] M. Wang and K. Lee, "FPGA-Based RISC-V for AI Acceleration in Surveillance Systems," IEEE Transactions on Embedded AI, 2024.

[5] J. Salamanca, R. Esteban, and L. F. Gonzalez, "A Dataset for Environmental Sound Recognition in Embedded Systems for Autonomous Vehicles," ResearchGate, 2024.

[6] K. Yang, J. Luo, F. Sun, and S. Wang, "Machine Learning for Multi-Sensor Fusion," Biased Algorithms, 2023.

[7] A. Mittal, "A Survey of Techniques for Implementing and Optimizing CNN Algorithms on FPGA," in Proc. NCAA, 2018.

[8] RISC-V International, "RISC-V for AI," IEEE Embedded Systems Conference, 2024.

[9] P. Culbertson, "Building Real-Time, Adaptive Robots Using Online Search," Cornell University, Dept. of Computer Science, Seminar, 2025.

[10] Y. Chen, J. Han, and L. Shao, "Multimodal Learning with Deep Boltzmann Machines," Pattern Recognition, 2017.

[11] N. Srivastava and R. R. Salakhutdinov, "Multimodal Learning with Deep Belief Nets," Journal of Machine Learning Research, 2012.

[12] G. P. Joshi and M. Kim, "A Survey on Sensor Fusion Techniques in Autonomous Vehicles," Sensors and Actuators, 2021.

[13] D. Calandra, A. Owens, D. Jayaraman, and S. Levine, "The Feeling of Success: Does Touch Sensing Help Predict Grasp Outcomes?," in Proc. Conf. on Robotics: Science and Systems (RSS), 2017.

[14] A. Sangiovanni-Vincentelli and G. Martin, "Platform-Based Design and Software Design Methodology for Embedded Systems," IEEE Design & Test of Computers, 2001.

[15] L. Zhang and F. Qian, "Review of FPGA-Based Acceleration for Deep Learning Models," in Proc. FPT, 2020.

[16] V. Govindaraju, P. Pannirselvam, and M. Ravichandran, "Design and Implementation of a RISC-V Processor on FPGA for IoT Applications," in Proc. ICCIDS, 2023.

[17] B. Lattner and C. Aditya, "Custom Instruction Design for RISC-V Embedded AI," in Proc. DAC, 2022.

[18] A. Rusu, M. Rabinowitz, and S. Blundell, "Progressive Neural Networks," arXiv:1606.04671, 2016.

[19] Y. Li, J. Liu, and H. Xu, "Edge Learning for Intelligent Control Systems," ACM Computing Surveys, 2022.

[20] M. Banerjee, R. Tiwari, and P. Mukherjee, "Efficient On-Device Learning for Edge Applications Using Reinforcement Strategies," in Proc. DATE, 2023.