Review of Role of AI in Embedded System

Kanawade Meera Vitthal
Department of Electronics & Telecommunication
Amrutvahini Polytechnic, Sangamner

Abstract: *The convergence of Artificial Intelligence (AI) and embedded systems has opened new possibilities for intelligent, real-time, and adaptive computing at the edge. This paper presents a comprehensive survey on the integration of AI in embedded systems, emphasizing architectures, algorithms, frameworks, and applications. We review key developments from academia and industry, discuss major challenges in resource-constrained environments, and highlight trends like TinyML, neuromorphic computing, and federated learning. The aim is to provide a consolidated reference for researchers and engineers working at the intersection of AI and embedded computing.*

Keywords: Artificial Intelligence, Embedded Systems, Edge AI, TinyML, Smart Devices, Real-Time Inference, Federated Learning

I. INTRODUCTION

Embedded systems are everywhere—from consumer electronics and industrial automation to smart healthcare and autonomous vehicles. Traditionally, these systems operated using rule-based logic and static configurations. With the advancement of AI, especially machine learning (ML) and deep learning (DL), these embedded devices are transforming into intelligent systems capable of perception, decision-making, and adaptation.AI offers embedded systems the ability to learn from data, recognize patterns, and respond to dynamic environments. This is critical in applications requiring low latency, offline operation, and privacy-preserving analytics. This survey explores the current state, ongoing research, and future directions of AI in embedded system.

Embedded System Overview

Embedded systems are computing platforms designed foe dedicated functions within larger mechanical or electrical systems. They are optimized for low power consumption ,Real time Performance and

Artificial Intelligence Fundamentals

AI involves computational techniques that enable machines to perform tasks that usually require human intelligence. These include learning, reasoning, and perception. With ML and DL, AI systems can now be trained on vast datasets and deployed on embedded hardware through model optimization.And this is a level 4 heading: It's recommended to write your text in a separate document and then add it to this template once it's complete. When copying text into the template from another document, make sure that the

II. LITERATURE SURVEY

Artificial Intelligence has increasingly found a home in embedded systems, thanks to advances in both hardware and software. Over the past decade, researchers have worked to adapt powerful AI techniques particularly machine learning and deep learningfor devices with limited memory, energy, and processing power

A. Making AI Models Smaller and Smarter

One of the earliest challenges was that AI models, especially deep neural networks, were simply too large to run on embedded hardware. Han et al. [2] addressed this problem with **Deep Compression**, a technique that significantly reduced the size of deep learning models using pruning, quantization, and Huffman coding—without a big loss in accuracy. Taking this idea further, the TinyML community emerged. Pioneered by Warden and Situnayake [3], TinyML focuses on deploying machine learning models on **ultra-low-power devices**, often with less than 1 MB of memory. Their work laid the foundation for today's edge AI solutions in wearables and IoT nodes.

B. Hardware to Support Embedded AI

To run AI efficiently, specialized hardware became essential. Chen et al. [4] provided an extensive review of **AI-enabled embedded platforms** like NVIDIA Jetson Nano, Google Coral, and Intel Movidius. These devices come with dedicated accelerators that handle neural network inference in real time. Their survey highlighted how hardware innovation is critical for enabling edge AI in areas like robotics and smart surveillance.

C. Learning Without the Cloud: Federated and On-Device Training

While many AI systems rely on cloud-based training, embedded systems often can't afford that luxury. To address privacy and bandwidth concerns, researchers like Li et al. [4] explored **federated learning**a method where AI models are trained locally on devices without sending raw data to a server.

D. Embedded AI in Real-World Applications

Beyond technical innovations, several works demonstrate how embedded AI is already being used in the real world. For instance, Sze et al. [5] presented a detailed tutorial on optimizing deep learning for embedded vision, showing how AI can be used for object detection in autonomous vehicles. Similarly, Kulkarni et al. [6] introduced an AI-powered smart agriculture system that uses embedded devices to monitor soil and crop health, helping farmers make better decisions.

III. AI TECHNIQUES FOR EMBEDDED SYSTEMS

A. Model Optimization Methods

1. Quantization: Converts weights and activations from 32-bit floats to 8-bit integers or smaller.quantization is the compression of floating-point data bits in neural network parameters to reduce model complexity and size by reducing the number of bits used by floating-point numbers, while maintaining model accuracy as much as possible.

2. Pruning: Removes unnecessary weights and neurons to reduce size and complexity.Pruning is a method used to reduce redundant data in the neural network by deter

mining the importance of each unit and removing unimportant parts.

Knowledge Distillation: Trains a lightweight model (student) to replicate a larger model (teacher).

B. Lightweight Frameworks and Tools

1. TensorFlow Lite-TensorFlow Mobile Reffered as LiteRT(Lite Runtime).It is designed to address the constraints of on device machine learning(ODML) such as privacy ,latency,connectivity,size and power consumption.liteRT supports running machine learning models directly on mobile devices,embedded system and microcontrollers without requiring server communication

2. PyTorch Mobile -PyTorch is a popular open-source machine learning framework that is increasingly being used in embedded systems. Its flexibility and ease of use, combined with its ability to perform deep learning tasks, make it a valuable tool for developing intelligent embedded applications.

3. CMSIS-NN (for ARM Cortex-M)-**CMSIS-N**N empowers developers to bring the capabilities of machine learning to a wide range of embedded systems by providing a highly optimized and efficient way to run neural networks on resource-constrained Arm Cortex-M processors.

4. Edge Impulse**-**Edge Impulse makes it easier for developers to integrate intelligence into embedded systems, enabling them to create smarter and more responsive devices

IV. APPLICATIONS

1. **Healthcare**-AI in embedded systems is revolutionizing healthcare by enabling more efficient, accurate, and personalized patient care. These systems can analyze real-time data, predict health issues, and automate tasks, leading to improved diagnostics, remote patient monitoring, and optimized treatment.AI powered embedded

systems analyze patient data (vital signs, medical images) for diagnosis, treatment planning, and patient monitoring, leading to more personalized and effectivehealthcare. Examples include AI-powered insulin pumps for dosageadjustments and medical imaging analysis.

2. **Automotive:** Self-driving cars use Artificial Intelligence(AI) to navigate and avoid accidents.
3. **Home Automatio**n: Devices like smart speakers, security cameras, and thermostats use AI for better control and comfort.
4. **Agriculture**-Smart farming tools use AI to monitor soil quality, control irrigation, and improve crop production.
5. **Smart Homes**-Embedded AI enables devices to learn user preferences (temperature, lighting) and automate tasks, improving energy efficiency and creating a more comfortable environment. This includes voice assistants that can understand and respond to user commands.
6. **Industrial Automation:** AI-powered embedded systems monitor equipment, perform predictive maintenance, and optimize production processes in smart factories. Intelligent robots can handle complex tasks with greater efficiency
7. **Consumer Electronics:** AI enhances the functionality of smartphones, smartwatches, and other devices through features like voice recognition, image processing, and predictive text, making them more user-friendly.
8. **Internet of Things (IoT)-**AI enhances IoT devices by enabling them to analyze data and make intelligent decisions locally

V. CHALLENGES

- **Limited Resources**: Embedded systems often have <1MB RAM and constrained CPU cycles.Embedded devices frequently possess restricted processing capabilities, memory capacity, and energy availability
- **Energy Consumption**: Power-efficient inference is essential for battery-powered devices.
- **Security and Privacy**: Sensitive data processed locally must be protected.Protecting embedded systems from hacking and security breaches is critical
- **Model Portability**: Deploying models across various hardware platforms is still challenging.

VI. EMERGING TRENDS

- **TinyML**: Running AI on microcontrollers with sub-milliwatt power budgets.
- **Neuromorphic Computing**: Brain-inspired chips like Intel Loihi for energy-efficient computation.
- **Federated Learning**: Training distributed AI models while preserving user privacy.
- **AI for System Design**: Neural Architecture Search (NAS) and co-design strategies for AI hardware/software

VII. CONCLUSION

The role of AI in embedded systems is becoming increasingly significant as we demand more intelligence from everyday devices. Through a combination of lightweight modeling, efficient hardware, and edge-focused learning techniques, AI is moving from data centers to the devices we carry and interact with. While challenges remain, the trajectory of research and innovation points toward an intelligent edge future.

REFERENCES

[1]. Zhaoyun Zhang and Jingpeng Li “Review of Artificial Intelligence in Embedded system” micromachines 2023,14(5),

[2]. Han, S., Mao, H., & Dally, W. J. (2016). *Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding*. ICLRI.

[3]. Dick, R.P.; Shang, L.; Wolf, M.; Yang, S.-W. Embedded Intelligence in the Internet-of-Things. IEEE Des. Test 2019, 37, 7–27. [Google Scholar] [CrossRef]

DOI: 10.48175/IJARSCT-28133

ISSN 2581-9429 IJARSCT

[4]. Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE, 105(12), 2295-2329.

[5]. Chen, T., Moreau, T., Jiang, Z., et al. (2019). *TVM: An automated end-to-end optimizing compiler for deep learning*. OSDI.

[6]. Banbury, C., Zhou, C., Fedorov, I., et al. (2020). *MLPerf Tiny Benchmark: Measuring the Performance of TinyML Systems*. arXiv:2005.08559.

[7]. Kulkarni, A., Kotecha, J., & Jain, M. (2021). *TinyML: Enabling AI in Embedded Systems*. IEEE Consumer Electronics Magazine, 10(2), 23-29.

.

DOI: 10.48175/IJARSCT-28133

ISSN 2581-9429 IJARSCT