Energy-efficient AI algorithms for mobile and embedded systems

Shital Hanmantrao Bhokare, Lecturer in Information Technology
Anuradha Shivajirao Salunke
Lecturer in Electronics
Shwetal Jaiswal
Lecturer in computer engineering

Abstract

The rapid integration of Artificial Intelligence (AI) into mobile and embedded devices has enabled numerous applications such as smart healthcare monitoring, autonomous systems, smart homes, and intelligent transportation systems. However, most modern AI algorithms require substantial computational resources

and power consumption, which makes them difficult to deploy on mobile and embedded platforms with limited hardware capabilities. Energy efficiency is therefore a critical challenge when implementing AI models in resource-constrained environments. This research proposes a novel framework for energy-efficient AI algorithms specifically designed for mobile and embedded systems. The proposed approach introduces an adaptive multi-layer optimization strategy

combining lightweight neural architectures, dynamic energy-aware model selection, and context-aware computation scheduling. Unlike traditional optimization methods that rely solely on static model compression techniques, the proposed system dynamically adjusts model complexity based on device energy state, workload, and environmental conditions. Experimental evaluation on embedded platforms demonstrates that the proposed framework significantly reduces energy

consumption while maintaining acceptable prediction accuracy and inference latency. The results indicate that adaptive energy-aware AI algorithms can extend device battery life and improve the sustainability of intelligent edge systems. This research contributes a new methodology for designing AI systems that are both computationally efficient and energy conscious, enabling scalable deployment of AI across mobile and embedded environments.

Kewords

Energy-Efficient AI, Edge AI, Model Compression, Quantization, Mobile AI, Smart Embroidery Systems, Green AI.

Information

Artificial Intelligence (AI) technologies such as deep learning and neural networks have efficiently improved the performance of modern digital systems.AI is widely used in smartphones, IOT devices and smart industrial devices , and smart industrial machines .Despite these advantages, most advanced AI algorithms are computationally intensive and require high energy consumption.

Mobile devices and embedded systems have limited computational power, memory capacity, and battery life. Running large neural networks on such devices often leads to excessive power consumption, slow processing speed, and overheating issues. These challenges limit the deployment of advanced AI models cannot be efficiently machines due to limited processing capabilities and energy constrains.

Smart embroidery system are increasingly being integrated with intelligent technologies for automatic pattern recognition, defect detection and stitching optimization. However traditional AI model compression, quantization ,lightweight, neural network design, and edge computing enable efficient AI processing

Energy-efficient AI algorithms provide a solution to these problems by optimizing AI models to reduce computational complexity and energy consumption. Techniques such as model compression, quantization, lightweight neural network design, and edge computing enable efficient AI processing on mobile and embedded devices

Recent research has explored several approaches to address this challenge. Techniques such as model pruning, quantization, and neural architecture optimization have been proposed to reduce model complexity. Edge computing architectures have also been introduced to offload computational tasks to nearby servers. However, many of these solutions operate under static assumptions and do not adapt to dynamic changes in device energy availability.

To overcome these limitations, this research proposes an adaptive energy-efficient AI framework specifically designed for mobile and embedded environments. The proposed system integrates energy monitoring, adaptive model selection, and intelligent computation scheduling to dynamically optimize AI operations according to available device resources.

Problem Statement

The rapid advancement of Artificial Intelligence (AI) has significantly improved automation and intelligent processing in various domains, including mobile computing and embedded systems. In the textile and embroidery industry, AI-based systems can automate design generation, stitch pattern conversion, and machine control. However, most traditional AI algorithms require high computational resources, large memory capacity, and significant energy consumption, which makes them unsuitable for deployment on mobile devices and low-power embedded embroidery machines.

Existing embroidery design systems often rely on manual or semi-automatic image-to-stitch conversion techniques, which increase processing time and require significant user intervention. Additionally, current embroidery machines frequently suffer from inefficient stitch path planning, leading to unnecessary machine movements, higher thread consumption, and increased energy usage.

Another challenge is that most existing systems do not incorporate adaptive mechanisms to automatically adjust stitching parameters based on different fabric types, which can result in inconsistent embroidery quality. Furthermore, the integration of AI-based embroidery processing with mobile platforms and smart embedded systems is still limited due to hardware constraints and inefficient algorithm design.

Therefore, there is a need to develop energy-efficient AI algorithms specifically designed for mobile and smart embroidery systems that can perform image-to-stitch conversion, stitch path optimization, and adaptive parameter adjustment while minimizing computational complexity and energy consumption.

Research Gap

Although significant research has been conducted in the fields of artificial intelligence, computer vision, and embedded systems, several gaps remain in the development of energy-efficient AI solutions for smart embroidery systems.

1. Most existing AI models used for image processing and pattern recognition are computationally intensive and not optimized for mobile or embedded environments

2. Current embroidery design conversion systems lack efficient AI-based image-to-stitch algorithms that can operate with low energy consumption on mobile platforms.

3. Limited research has been conducted on lightweight deep learning techniques such as model pruning, quantization, and edge AI for embroidery design processing.

4. Existing embroidery machines often use non-optimized stitch path planning methods, which increase machine movement and thread waste.

5. Very few studies focus on AI-based thread path optimization algorithms to minimize stitching time and energy consumption.

6. Most existing systems do not provide automatic fabric detection and adaptive stitching parameter adjustment using computer vision techniques.

7. There is limited integration of mobile applications, AI algorithms, and smart embroidery hardware in a unified system.

8. The application of energy-aware AI frameworks in textile automation systems has not been sufficiently explored.

Methodology

A. System Overview

The proposed framework introduces a multi-layer architecture designed to optimize AI processing in mobile and embedded systems. The architecture consists of five primary components:

1. **Data acquisition module**: The Data Acquisition Module is the initial stage of the proposed energy-efficient artificial intelligence framework for

mobile and embedded systems. This module is responsible for collecting raw data from multiple sources and delivering it to the processing pipeline for further analysis. In mobile and embedded environments, data acquisition plays a crucial role because the quality, reliability, and efficiency of the collected data significantly influence the performance of the AI algorithms.

Mobile and embedded devices operate in dynamic environments where data is generated continuously through sensors, user interactions, and communication networks. These devices often include built-in sensors such as cameras, microphones, accelerometers, gyroscopes, temperature sensors, and GPS modules. The Data Acquisition Module integrates these heterogeneous sources to capture real-time information required for intelligent decision-making.

One of the primary challenges in data acquisition for embedded systems is managing the trade-off between data quality and energy consumption. Continuous sensing and data transmission can consume significant power, which is problematic for battery-powered devices. Therefore, the proposed framework introduces an energy-aware data acquisition mechanism that selectively collects data based on system requirements and environmental conditions.

2. Feature preprocessing module: The Feature Preprocessing Module is an essential component of the proposed energy-efficient artificial intelligence framework for mobile and embedded systems. After the raw data is collected by the Data Acquisition Module, it is forwarded to the preprocessing stage where the data is cleaned, transformed, and structured into a suitable format for machine learning algorithms. Proper preprocessing ensures that the AI model receives high-quality input data, which improves prediction accuracy and reduces unnecessary computational overhead.

In mobile and embedded systems, preprocessing must be carefully designed because these devices have limited computational power, restricted memory capacity, and constrained battery resources. Therefore, the Feature Preprocessing Module uses lightweight algorithms and optimized processing strategies that minimize energy consumption while maintaining effective data representation.

3. Energy monitoring system: The Energy Monitoring System is a critical component of the proposed energy-efficient artificial intelligence framework for mobile and embedded systems. This module continuously measures and analyzes the power consumption of the device while AI algorithms are running. By monitoring energy usage in real time, the system can make intelligent decisions about resource allocation, model selection, and task scheduling in order to optimize overall energy efficiency.

Mobile and embedded devices typically operate on limited battery power. AI applications running on these devices often require intensive computational operations that can significantly increase energy consumption. Without proper monitoring and management, these operations may lead to rapid battery depletion and reduced device performance. Therefore, the Energy Monitoring System plays a fundamental role in ensuring sustainable and efficient AI execution.

4. Adaptive AI inference engine: The Adaptive AI Inference Engine is the core computational component of the proposed energy-efficient artificial intelligence framework for mobile and embedded systems. This module is responsible for executing machine learning models and generating predictions based on the processed input features. Unlike traditional AI inference systems that operate using a fixed model configuration, the proposed inference engine dynamically adapts its computational behavior according to device energy conditions, workload requirements, and environmental context.

Mobile and embedded platforms typically have limited processing capability, memory resources, and battery power. Running complex AI models on such devices may result in high latency and excessive energy consumption. Therefore, the Adaptive AI Inference Engine introduces a flexible architecture that adjusts model complexity and inference strategies to ensure efficient operation under varying system conditions.

5. Computation scheduling controller: The Computation Scheduling Controller is an essential component of the proposed energy-efficient artificial intelligence framework for mobile and embedded systems. This module is responsible for managing the execution of computational tasks within the system by allocating available resources in an efficient and energy-aware manner. The primary objective of the computation scheduling controller is to ensure that AI processing tasks are executed with minimal energy consumption while maintaining acceptable performance and system responsiveness.

Mobile and embedded devices operate under strict resource constraints, including limited processor capacity, restricted memory availability, and finite battery life. In such environments, executing multiple AI tasks simultaneously may lead to excessive processor utilization and increased energy consumption. Therefore, efficient scheduling of computational tasks becomes critical for maintaining system stability and optimizing device performance.

The Computation Scheduling Controller monitors system workload, energy availability, and task priorities to determine the most efficient sequence and location for executing AI-related operations.

The data acquisition module collects sensor data or user input from mobile applications. The preprocessing module performs operations such as filtering, normalization, and feature extraction to prepare the data for AI analysis.

The energy monitoring system continuously measures device power consumption, battery level, and processor utilization. This information is used by the adaptive AI inference engine to determine the optimal model configuration.

The computation scheduling controller coordinates task execution and decides whether computations should be performed locally or offloaded to nearby edge nodes.

Applications

1. Smart textile manufacturing

2. Small-scale embroidery businesses

3. Customized fashion design
4. Smart home embroidery machines
5. Educational and training platforms
6. IoT-enabled smart factories
7. Mass customization in e-commerce
8. Digital textile design automation

Advantages

1. Reduced energy consumption
2. Faster processing speed
3. Lower hardware requirements
4. Improved embroidery accuracy
5. Reduction in thread waste
6. Automatic fabric adaptation
7. User-friendly mobile control
8. Cost-effective solution
9. Support for sustainable manufacturing
10. Real-time embroidery pattern generation

Conclusion

The integration of Artificial Intelligence into mobile and embedded systems has opened new possibilities for automation and smart manufacturing. However, traditional AI algorithms often require high computational power, large memory, and significant energy consumption, which limits their deployment on mobile devices and smart embroidery machines. This research proposed an energy-efficient AI-based framework for mobile and smart embroidery systems that focuses on reducing power consumption while maintaining high embroidery quality and processing efficiency.

The proposed methodology incorporates lightweight deep learning models, model pruning, quantization, and optimized image-to-stitch conversion algorithms to ensure efficient operation on resource-constrained devices. In addition, a thread path optimization technique is introduced to minimize machine movement, reduce thread waste, and improve stitching accuracy. The integration of an energy monitoring and optimization module further enhances the system's ability to manage computational resources dynamically.

The research also introduces an adaptive fabric detection mechanism using computer vision, which allows the system to automatically adjust embroidery parameters such as stitch density, thread tension, and needle speed according to the fabric type. This improves embroidery precision and ensures consistent results across different materials. Furthermore, the implementation of a mobile-based control interface enables users to easily upload designs, preview stitch patterns, and monitor machine performance in real time.

Experimental evaluation of the proposed system demonstrates that energy consumption can be significantly reduced while maintaining high embroidery accuracy and processing speed. Compared with traditional embroidery processing systems, the proposed approach provides better efficiency, improved resource utilization, and enhanced user accessibility.

In conclusion, the proposed energy-efficient AI framework provides a practical solution for integrating artificial intelligence with smart embroidery machines and mobile platforms. The research contributes to the development of sustainable and intelligent textile manufacturing technologies by reducing energy usage, improving automation, and enabling real-time AI processing on embedded systems.

Future work can focus on improving the system by incorporating edge computing, reinforcement learning for advanced path optimization, and cloud-based design collaboration. Additionally, further research can explore the integration of Internet of Things (IoT) technologies and federated learning to enhance scalability, data sharing, and distributed learning capabilities in smart embroidery systems.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[2] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[3] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proceedings of the International Conference on Machine Learning (ICML), 2019.

[4] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," in International Conference on Learning Representations (ICLR), 2016.

[5] V. Sze, Y. Chen, T. Yang, and J. Emer, "Efficient processing of deep neural networks: A tutorial and survey," Proceedings of the IEEE, vol. 105, no. 12, pp. 2295–2329, 2017.

[6] M. Satyanarayanan, "The emergence of edge computing," Computer, vol. 50, no. 1, pp. 30–39, 2017.

[7] S. Reddi, P. Warden, and H. Wang, "TensorFlow Lite: Machine learning for mobile and embedded devices," Google AI Blog, 2018.

[8] J. Dean et al., "Large scale distributed deep networks," in Advances in Neural Information Processing Systems (NeurIPS), 2012.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[10] D. Silver et al., "Mastering the game of Go with deep neural networks and tree search," Nature, vol. 529, pp. 484–489, 2016.