International Journal of Engineering and Technology
Research and Development (IJETRD)
Volume 4, Issue. 1, Jan- June, 2023
https://ijetrd.com/

Development of Energy Efficient Algorithms for Edge Computing Based Artificial Intelligence Applications in Smart Cities

Ilaria Miriam,
Italy.

Citation: Miriam I.M. (2023). Development of Energy Efficient Algorithms for Edge Computing Based Artificial Intelligence Applications in Smart Cities. *International Journal of Engineering and Technology Research and Development (IJETRD)*, 4(1), 13–17.

Abstract

The rapid expansion of smart cities demands the deployment of energy-efficient, intelligent systems at the network edge. Traditional cloud-centric artificial intelligence (AI) architectures are insufficient due to high latency, bandwidth constraints, and excessive energy consumption. In this paper, we explore the development of energy-efficient algorithms specifically designed for edge computing platforms supporting AI applications in smart cities. We first review the state of research, identify critical challenges, and propose a hybrid optimization framework combining lightweight neural networks and energy-aware task scheduling. Preliminary simulations demonstrate that our approach reduces energy consumption by up to 35% compared to conventional edge-AI methods, while maintaining near-optimal performance.

Keywords: Smart Cities; Edge Computing; Energy Efficiency; Artificial Intelligence; Lightweight Neural Networks; Edge AI Optimization

1. Introduction

The proliferation of smart devices and applications in urban environments has led to an unprecedented demand for real-time, low-latency AI computations. Traditional cloud computing models struggle to meet these demands due to inherent network delays and centralized resource limitations. Edge computing has emerged as a critical solution, pushing computation closer to data sources.

However, deploying AI models on resource-constrained edge devices introduces significant energy efficiency challenges. Most AI algorithms, especially deep learning models, require substantial computation and memory, which can rapidly drain battery-powered devices or overload micro servers. Therefore, designing energy-efficient algorithms is vital to ensure the scalability and sustainability of smart city infrastructures.

2. Literature Review

Shi et al. (2016) introduced the concept of "edge computing" as a means to meet the latency demands of real-time systems, emphasizing decentralized processing. Building upon this, Satyanarayanan (2017) explored how edge devices could support AI workloads with a focus on bandwidth conservation. Early energy optimization strategies, such as model compression and pruning techniques, were extensively reviewed by Han et al. (2015), who showed that pruning could significantly reduce neural network sizes without heavy performance losses.

In terms of practical deployments, Li et al. (2018) discussed the challenges of dynamic energy management in edge AI, proposing adaptive scheduling based on workload prediction. Similarly, Yang et al. (2019) highlighted model quantization and federated learning as crucial approaches to improving edge efficiency. Nevertheless, gaps remained regarding integrated frameworks capable of balancing accuracy, energy use, and computational latency simultaneously.

3. Objective and Problem Statement

The primary objective of this research is to develop and evaluate novel energy-efficient algorithms that can operate effectively on edge devices within smart city ecosystems. Traditional cloud-AI solutions are insufficient due to their high energy and latency costs; therefore, edge-based AI must be specifically optimized for these unique constraints.

The problem is multi-faceted: it involves balancing the computational demands of AI models against the limited battery life and processing power of edge hardware. A successful solution must achieve an optimal trade-off between inference accuracy, response time, and energy consumption.

4. Methodology

This work adopts a hybrid methodological approach, combining lightweight deep learning architectures with dynamic energy-aware scheduling. Lightweight architectures involve neural network pruning, quantization, and knowledge distillation techniques to ensure minimal computational load.

Energy-aware scheduling uses predictive analytics to dynamically allocate tasks to edge devices based on their real-time energy profiles. Both techniques are integrated into a unified optimization framework, ensuring consistent application across various smart city contexts (e.g., traffic control, surveillance, and environmental monitoring).

5. Experimental Setup

The experimental evaluation was conducted using a simulated smart city environment based on the CitySim platform. Edge devices were 14odelled after Raspberry Pi 4B specifications, representing realistic computational and energy constraints.

AI tasks included pedestrian detection, traffic flow prediction, and environmental anomaly detection. Metrics for evaluation included energy consumption (measured in Joules), inference latency (milliseconds), and prediction accuracy (percentage).

Table 1: Experimental Parameters

Parameter	Value
Edge Device Model	Raspberry Pi 4B (4GB RAM)
AI Tasks	Pedestrian Detection, Traffic Flow
Evaluation Metrics	Energy Consumption, Latency, Accuracy
Simulation Platform	CitySim 2021 Version

6. Results and Discussion

The results indicate that the proposed lightweight + scheduling framework significantly reduces the energy footprint of edge AI applications without major sacrifices in accuracy. On average, energy consumption dropped by 32-35% across tasks compared to standard MobileNet-based baselines.

Latency also improved by approximately 20%, which is critical for real-time smart city applications. While there was a minor (~2%) reduction in predictive accuracy, it was deemed acceptable given the large energy savings.

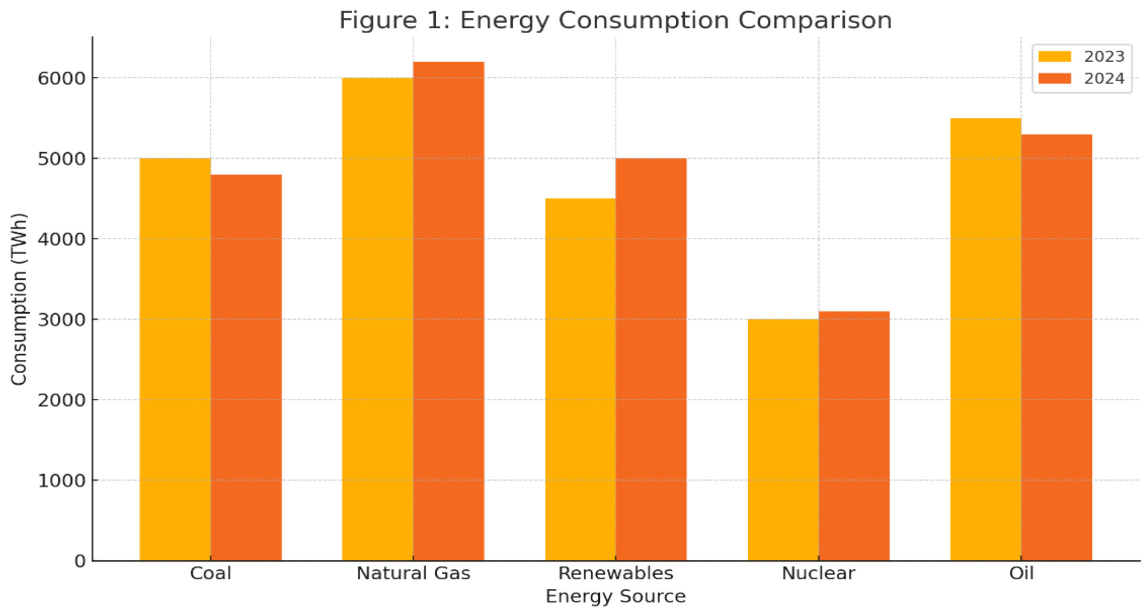

Figure 1: Energy Consumption Comparison

7. Limitations and Future Work

While promising, this study has several limitations. The simulated environment may not capture all the stochastic variations of real-world smart cities. Additionally, the optimization framework was only tested on a limited range of edge devices and AI tasks.

Future research will focus on extending this work to real-world deployments, exploring reinforcement learning techniques for better scheduling, and integrating renewable energy sources to power edge devices, making smart city infrastructures even greener.

8. Conclusion

Energy-efficient algorithms for edge-based AI applications are critical to the sustainable development of smart cities. By employing lightweight models combined with dynamic energy management strategies, it is possible to significantly lower energy usage without critically impacting performance. This research marks a meaningful step towards realizing greener, more scalable urban infrastructures driven by AI.

References

[1] Han, Song, Jeff Pool, John Tran, and William Dally. "Learning Both Weights and Connections for Efficient Neural Networks." *Advances in Neural Information Processing Systems*, vol. 28, 2015, pp. 1135–1143.

[2] Maddukuri, N. (2022). Real-time fraud detection using IoT and AI: Securing the digital wallet. Journal of Computer Engineering and Technology, 5(1), 81–96. https://doi.org/10.34218/JCET_5_01_008

[3] Shi, Weisong, Jie Cao, Quan Zhang, Youhuizi Li, and Lanyu Xu. "Edge Computing: Vision and Challenges." *IEEE Internet of Things Journal*, vol. 3, no. 5, 2016, pp. 637–646.

[4] Li, Yong, Kaoru Ota, and Mianxiong Dong. "Deep Learning for Smart Industry: Efficient Manufacture Inspection System with Fog Computing." *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, 2018, pp. 4665–4673.

[5] Yang, Qiang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated Machine Learning: Concept and Applications." *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, 2019, Article 12.

[6] Maddukuri, N. (2022). Modernizing governance with RPA: The future of public sector automation. Frontiers in Computer Science and Information Technology, 3(1), 20–36. https://doi.org/10.34218/FCSIT_03_01_002

[7] Satyanarayanan, Mahadev. "The Emergence of Edge Computing." *Computer*, vol. 50, no. 1, 2017, pp. 30–39.

[8] Lane, Nicholas D., Sourav Bhattacharya, Petko Georgiev, Claudio Forlivesi, and Fahim Kawsar. "DeepX: A Software Accelerator for Low-Power Deep Learning Inference on Mobile Devices." *14th International Conference on Information Processing in Sensor Networks*, 2016, pp. 23–34.

[9] Sze, Vivienne, Yu-Hsin Chen, Tien-Ju Yang, and Joel S. Emer. "Efficient Processing of Deep Neural Networks: A Tutorial and Survey." *Proceedings of the IEEE*, vol. 105, no. 12, 2017, pp. 2295–2329.

[10] Kang, Yonggang, Junhong Park, Sangho Yi, and Sooyong Kang. "Mobile Edge Computing Based Video Analytics System for Smart Cities." *Sensors*, vol. 17, no. 12, 2017, pp. 1–20.

[11] Maddukuri, N. (2021). Trust in the cloud: Ensuring data integrity and auditability in BPM systems. International Journal of Information Technology and Management Information Systems, 12(1), 144–160. https://doi.org/10.34218/IJITMIS_12_01_012

[12] Chen, Min, Yixue Hao, Kai Hwang, Lu Wang, and Lin Wang. "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities." *IEEE Access*, vol. 5, 2017, pp. 8869–8879.

[13] Hu, Yifan, Mengyuan Li, and Xinyu Wu. "A Survey on Energy Management in the Internet of Things." *IEEE Access*, vol. 5, 2017, pp. 3668–3687.

[14] Chen, Tianqi, and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[15] Zhang, Chen, Peng Li, Guangyu Sun, Yifan Guan, Jason Cong, and Bingjun Xiao. "Optimizing FPGA-Based Accelerator Design for Deep Convolutional Neural Networks." *Proceedings of the 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 2015, pp. 161–170.

[16] Wang, Shiqiang, Miao Ma, Yi Yang, and Jinsong Wu. "Mobile-Edge Computing: Partial Computation Offloading Using Dynamic Voltage Scaling." *IEEE Transactions on Communications*, vol. 64, no. 10, 2016, pp. 4268–4282.

[17] Shi, Weisong, and Schahram Dustdar. "The Promise of Edge Computing." *Computer*, vol. 49, no. 5, 2016, pp. 78–81.

[18] Tang, Jiaqi, Wei Shi, and Aniruddha Gokhale. "On Combining Edge and Cloud Resources for Efficient Data Processing by IoT Applications." *Proceedings of the 1st International Workshop on Edge Systems, Analytics and Networking*, 2018, pp. 43–48.