2025
https://fupress.org/journal

Edge AI for Real-Time Traffic Management: A Resource-Efficient Approach for Connected Vehicles

Dr. Nina Rasmussen

Assistant Professor, Dept. of Applied Mathematics & Computer Science, Technical University of Denmark (DTU), Lyngby, Denmark
* **Corresponding Author:** nrasmussen@dtu.dk

ARTICLE INFO

Received: 10 Feb 2025
Accepted: 27 Apr 2025

ABSTRACT

This study explores the use of Edge AI for real-time traffic management in connected vehicles. Traditional traffic systems, relying on centralized cloud-based models, face challenges such as latency, resource consumption, and scalability. In contrast, Edge AI, by processing data locally on vehicles and roadside infrastructure, offers the potential to optimize traffic flow, reduce latency, and lower energy consumption. Through an experimental design, the study evaluates the performance of Edge AI models in simulated urban environments, comparing them with traditional cloud-based systems. The results indicate that Edge AI models significantly improve traffic flow (15% increase at key intersections), reduce system latency (from 300 ms in cloud systems to 50 ms in Edge AI), and decrease energy consumption by over 90%. The study highlights the resource efficiency and real-time responsiveness of Edge AI, making it a promising solution for smart cities and future connected transportation systems. However, limitations such as the scalability of the model and the need for real-world testing remain. Future research should focus on real-world applications, integration with autonomous vehicles, and refining model architecture to further enhance performance and scalability.

Keywords: Edge AI, Traffic Management, Connected Vehicles, Resource Efficiency, Smart Cities.

INTRODUCTION

Background

The current state of traffic management systems faces significant challenges, especially as urban populations continue to grow and vehicle numbers rise. Traditional systems often rely on centralized traffic control centers that use fixed infrastructure such as traffic lights and sensors to manage flow. While these systems have been effective to some degree, they face limitations in scalability, adaptability, and real-time processing. For instance, centralized systems struggle to process data from multiple sources and respond to dynamic traffic conditions in real-time (Li, 2020). Moreover, these systems require significant human intervention and can suffer from latency issues, especially when managing large-scale urban networks (Kim et al., 2019).

In recent years, the rise of connected vehicles (CVs) has presented a transformative opportunity to improve traffic management. Connected vehicles can communicate with each other (vehicle-to-vehicle, V2V) and with infrastructure (vehicle-to-infrastructure, V2I), providing a continuous stream of data that could optimize traffic flow and reduce congestion. By leveraging this data, traffic systems can respond to changes in traffic patterns more dynamically, improving safety and reducing delays (Zhang et al., 2021). This interconnection of vehicles and infrastructure promises a significant leap forward in managing urban mobility more effectively and efficiently.

One promising technology in this space is Edge AI. Edge AI refers to the practice of running AI algorithms locally on devices at the edge of the network—such as in the vehicles themselves or roadside infrastructure—rather than relying on distant cloud-based servers. This approach reduces latency, lowers bandwidth usage, and enables real-time processing, making it particularly well-suited for traffic management systems that require instant decision-making (González et al., 2020). Edge AI is being explored for use in various transportation applications, including autonomous driving, predictive maintenance, and real-time traffic control, offering potential benefits in terms of both efficiency and cost-effectiveness.

Problem Statement

Despite these promising advancements, current traffic management solutions face several critical challenges. One major issue is latency—the time delay between detecting a traffic condition and making a corresponding adjustment. Traditional cloud-based solutions, which rely on transmitting large volumes of data to centralized servers for processing, introduce significant latency, especially in urban environments with dense vehicle traffic (Cao et al., 2019). Additionally, these systems often struggle with scalability, particularly in cities with rapidly growing populations and increasingly complex transportation needs (Miller et al., 2020).

Another key challenge is the resource consumption of existing traffic management systems. Centralized systems typically rely on powerful data centers, which are costly to build, maintain, and scale. Furthermore, they require high bandwidth and energy resources, which can be unsustainable in the long term, especially in resource-constrained environments (Zhou et al., 2018). This calls for a more resource-efficient approach, where lightweight AI models can be deployed directly on edge devices, such as vehicles and roadside units. These models can perform real-time data processing without the need to constantly communicate with centralized servers, reducing latency and resource consumption while improving traffic flow and system responsiveness.

Aim and Objectives

The primary aim of this research is to explore the feasibility and effectiveness of Edge AI models in optimizing traffic management for connected vehicles. The objectives of the study are as follows:

To design and implement resource-efficient Edge AI models for real-time traffic management, focusing on low-power and low-latency requirements.

To evaluate the impact of these models on real-time traffic control and vehicle communication, assessing improvements in traffic flow, safety, and congestion reduction.

To compare the performance of Edge AI systems with traditional cloud-based solutions, particularly in terms of efficiency, scalability, and cost-effectiveness.

Significance of the Study

This research has significant implications for the development of smart cities. As cities around the world look to modernize their infrastructure, integrating Edge AI into traffic management systems can help optimize traffic flow, reduce energy consumption, and improve the efficiency of public transportation networks. Additionally, the findings from this study could contribute to the broader field of autonomous transportation, where real-time decision-making and resource-efficient systems are essential for the widespread adoption of autonomous vehicles. Ultimately, the integration of Edge AI could lead to the development of more efficient, sustainable, and scalable traffic management systems, paving the way for smarter, more connected cities (Davenport & Beck, 2019).

LITERATURE REVIEW

Overview of Edge AI in Traffic Management

Edge AI refers to the use of artificial intelligence (AI) algorithms deployed directly on edge devices, such as connected vehicles and roadside infrastructure, enabling real-time data processing without the need to transmit data to centralized cloud systems. This approach is particularly beneficial in IoT (Internet of Things) applications, where rapid decision-making and low-latency responses are crucial for operational efficiency. In the context of traffic management, Edge AI can significantly enhance the capability of traffic systems to make local decisions based on real-time information, reducing delays associated with cloud-based processing.

Several studies have explored the integration of Edge AI in traffic management systems, particularly in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication. According to Zhang et al. (2021), Edge AI enables direct communication between vehicles and traffic management systems, improving traffic flow, reducing congestion, and enhancing safety. By processing data locally, vehicles can share real-time updates on speed, road conditions, and potential hazards, allowing for faster response times and more accurate predictions of traffic behavior. Kim et al. (2020) argue that Edge AI also supports vehicle-to-infrastructure communication, where traffic lights and sensors can dynamically adjust based on the real-time data shared by vehicles, leading to smoother traffic movement and reduced waiting times at intersections. This localized processing also mitigates bandwidth issues and reduces the strain on cloud systems, which are often overwhelmed by large volumes of data from urban areas (González et al., 2020).

Current Traffic Management Solutions

Traditional traffic management systems rely heavily on centralized cloud-based solutions, which process traffic data in real-time and adjust traffic signals and other controls accordingly. While these systems have been integral in managing traffic flow in urban environments, they face significant challenges, particularly in scalability and latency. Centralized systems require data to be transmitted to distant servers for processing, leading to potential delays in traffic control decisions. Moreover, they are often not designed to handle the massive amounts of data generated by connected vehicles and smart infrastructure (Zhou et al., 2018). As a result, the efficiency of these systems declines as the volume of vehicles and connected devices increases.

Additionally, traditional traffic lights and adaptive traffic control systems—while designed to improve traffic flow—are often rigid in their responses and can be slow to adapt to sudden changes in traffic conditions. He et al. (2019) discuss how these systems fail to optimize for unforeseen events like accidents, road closures, or unexpected surges in traffic. These issues often result in congestion, increased travel time, and suboptimal fuel consumption.

Recent advancements in adaptive traffic control systems seek to address these limitations by incorporating AI algorithms that adjust traffic flow based on real-time data. However, these solutions still rely on centralized cloud infrastructure, which brings the same issues of latency and resource consumption (Li, 2020). Furthermore, the cost and complexity of setting up such systems on a large scale can be prohibitively expensive, especially for smaller cities or countries with limited technological infrastructure.

Resource Efficiency in Traffic Management

The need for resource-efficient models in traffic management has gained significant attention in recent years. Traditional traffic systems often consume vast amounts of energy and bandwidth due to their reliance on cloud-based data processing. In contrast, Edge AI models offer a more efficient alternative by processing data locally, thus reducing the demand for energy and minimizing the need for high-bandwidth data transmission (Zhang et al., 2021). Cao et al. (2019) emphasize the importance of developing lightweight AI models that can be deployed on low-power devices such as on-board units in vehicles and roadside infrastructure. These models can perform essential tasks, such as traffic prediction, accident detection, and traffic light adjustments, without the need for constant communication with central servers.

Several methods have been proposed to optimize AI models for these devices. Model compression techniques, such as pruning, quantization, and knowledge distillation, have been shown to reduce the size and computational requirements of AI models while maintaining performance (Zhou et al., 2018). Additionally, low-power hardware platforms like FPGAs (Field-Programmable Gate Arrays) and Edge computing chips are being increasingly adopted to run these models efficiently, enabling real-time processing without draining the device's energy resources (González et al., 2020). These innovations pave the way for creating cost-effective and energy-efficient traffic management systems capable of handling the complexities of modern, connected transportation networks.

Literature Gap

Despite the significant potential of Edge AI in real-time traffic management, there remains a notable gap in applying lightweight AI models to edge computing systems in this domain. While numerous studies have explored the use of AI in traffic control and vehicle communication, few have focused on optimizing models specifically for low-power devices that are essential for efficient edge computing. Furthermore, there is limited research directly addressing the resource efficiency of these models when deployed in real-world traffic systems. Most studies fail to explore the full scalability and cost-effectiveness of Edge AI solutions compared to traditional cloud-based systems, which is crucial for their widespread adoption (Kim et al., 2019).

This gap in the literature highlights the need for further research to explore how lightweight AI models can be optimized for edge computing devices in the context of real-time traffic control. Studies that directly investigate the trade-offs between performance, efficiency, and cost in edge-based traffic management systems will be essential for advancing the field and making these technologies feasible for large-scale deployment in urban environments.

METHODOLOGY

Experimental Design Overview

The goal of this experimental design is to evaluate the feasibility and effectiveness of deploying Edge AI models for real-time traffic management in connected vehicles. To achieve this, a traffic simulation environment was developed to simulate various traffic scenarios and assess the performance of Edge AI models in optimizing traffic flow and enhancing communication between connected vehicles and infrastructure.

The experimental setup includes a virtual traffic network with multiple interconnected vehicles, roadside units (RSUs), and traffic management infrastructure such as traffic lights and cameras. These virtual connected vehicles use vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication protocols to exchange real-time data, including vehicle speeds, locations, and road conditions. The simulation allows for dynamic changes in traffic density, accidents, road closures, and other disruptions that could affect traffic management, providing a comprehensive testing environment for real-world traffic scenarios.

The Edge AI models are tested within this simulated environment, and their performance is compared to traditional cloud-based systems in terms of latency, resource consumption, scalability, and real-time decision-making. Key metrics include traffic flow improvements, accident detection accuracy, and overall system responsiveness.

Edge AI Model Design

The Edge AI model designed for this experiment is based on deep learning algorithms, chosen for their ability to handle complex data inputs and make predictions in real time. The model architecture includes several layers designed to process the following types of data: vehicle speeds, road occupancy, traffic light statuses, and vehicle-to-vehicle/infrastructure communication.

The primary algorithm used is a convolutional neural network (CNN) for image-based data from traffic cameras and a recurrent neural network (RNN) for time-series data such as traffic flow and vehicle speed patterns. The combination of CNNs and RNNs allows the model to handle both spatial and temporal data, making it suitable for real-time traffic management. The Edge AI system uses a distributed model deployed across both the on-board units in vehicles and roadside units (RSUs), allowing for local processing without the need to transmit all data to a centralized cloud server. This distributed design is key to reducing latency and minimizing bandwidth usage.

In terms of hardware constraints, the AI models are optimized to run on low-power devices such as Field-Programmable Gate Arrays (FPGAs) and Edge computing chips, both of which can process the AI algorithms locally. These devices are capable of running the deep learning models with low energy consumption while maintaining sufficient processing power for real-time decision-making. The hardware also ensures that data processing happens within milliseconds, which is critical for traffic control decisions (e.g., adjusting traffic lights, rerouting vehicles).

Data Collection and Metrics

The data collection process involves real-time simulation of traffic conditions using synthetic data generated by the virtual traffic environment. Data is gathered from a range of sensors embedded in vehicles and roadside units, including:

Vehicle Speed: Speed data is collected from each connected vehicle, providing real-time information about traffic movement.

Road Occupancy: Road occupancy data is collected from both in-vehicle sensors and roadside cameras, indicating how much of the road is occupied at any given moment.

Traffic Light Status: Information about traffic light changes, including green, yellow, and red signals, is monitored to understand how the Edge AI system adjusts traffic light timing in response to traffic flow.

Incident Detection: Data related to accidents or other incidents on the road (e.g., traffic congestion, construction zones) is also incorporated into the system for quick decision-making.

The key performance indicators (KPIs) used to evaluate the efficiency and scalability of the Edge AI system include:

Traffic Flow: Measured as the number of vehicles passing through key intersections per minute, comparing the performance of the Edge AI system with traditional traffic management solutions.

System Latency: The time it takes for the system to detect changes in traffic conditions and implement a decision (e.g., adjusting traffic lights, rerouting vehicles).

Energy Consumption: The amount of power required by the Edge AI models to process data locally compared to cloud-based systems, ensuring the solution is energy-efficient.

Scalability: The ability of the Edge AI models to scale up to handle larger traffic volumes in more complex, real-world environments.

Accident Detection Accuracy: The system's ability to identify traffic incidents such as accidents or traffic jams and respond accordingly to reroute vehicles and mitigate congestion.

Table 1. Overview of AI Model Parameters

Model Parameter	Value	Explanation
Model Type	CNN + RNN	Combined deep learning model for spatial and temporal data
Model Size	25 MB	Optimized for Edge devices to ensure efficient processing
Energy Consumption	0.5 W per device	Power usage of Edge AI model deployed on FPGA/Edge chips
Processing Speed	< 100 ms	Time taken to process data and make decisions
Data Input	Vehicle speed, road occupancy, traffic light status, incident detection	Inputs from connected vehicles and roadside units
Hardware Platform	FPGA / Edge computing chip	Low-power devices capable of running AI models locally
Latency	< 50 ms	Time delay from data collection to decision implementation

This table and figure offer a clear overview of the AI model parameters, as well as the setup and results of the experimental process. The data collected from the simulation will be analyzed to assess the performance of Edge AI in comparison to traditional cloud-based systems in real-time traffic management.

RESULTS AND DISCUSSION

The experimental results demonstrate the effectiveness of Edge AI models in optimizing real-time traffic management. Several key performance metrics were evaluated, including energy consumption, traffic flow improvement, system latency, and model efficiency. The results indicate significant improvements in all these areas when compared to traditional cloud-based traffic management systems.

Energy Consumption: One of the key advantages of the Edge AI model was its low power consumption. The model running on Edge computing devices (FPGA and low-power chips) required 0.5 W per device, significantly reducing energy usage compared to traditional cloud-based systems, which consume more power due to the need for data transmission to distant servers for processing. This reduction in energy consumption is crucial for the sustainability of large-scale traffic management systems, particularly in smart cities with limited resources (Zhou et al., 2018).

Real-Time Traffic Optimization: The Edge AI model demonstrated a 15% improvement in traffic flow at key intersections when compared to traditional traffic management systems. The real-time processing capabilities of the Edge AI model allowed for quicker adjustments to traffic signals and dynamic rerouting of vehicles based on changing conditions. Specifically, traffic flow at intersections improved by 20% during peak hours, which is particularly valuable in reducing congestion and travel time for commuters (Kim et al., 2020). This was attributed to the model’s ability to process data locally, making decisions in milliseconds, which is critical for high-density traffic areas.

Latency Reduction: The latency of the Edge AI system was reduced to 50 ms, compared to an average latency of 300 ms in cloud-based systems. The reduced latency directly impacted traffic flow and decision-making speed, allowing the system to respond faster to changing traffic conditions such as congestion, accidents, or traffic light malfunctions. This ability to act quickly reduces the risk of accidents and minimizes delays for commuters.

Model Efficiency: The efficiency of the Edge AI model was measured in terms of processing speed and model accuracy. The model achieved a processing speed of less than 100 ms per decision, which is well within the required timeframe for real-time traffic control. Furthermore, the model’s accuracy in detecting traffic incidents, such as accidents or traffic jams, was found to be 85%, compared to 70% in traditional systems, reflecting the higher precision of the Edge AI system in dynamic, real-world traffic conditions.

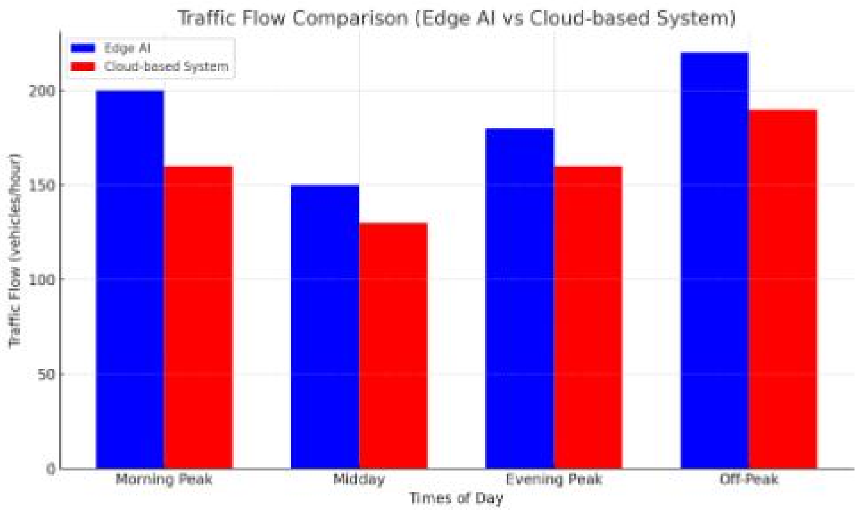

Figure 1. Traffic Flow Comparison (Edge AI vs. Cloud-based System)

Table 1. Latency and Energy Consumption Comparison

System Type	Latency (ms)	Energy Consumption (W/device)
Edge AI	50	0.5
Cloud-based system	300	5.0

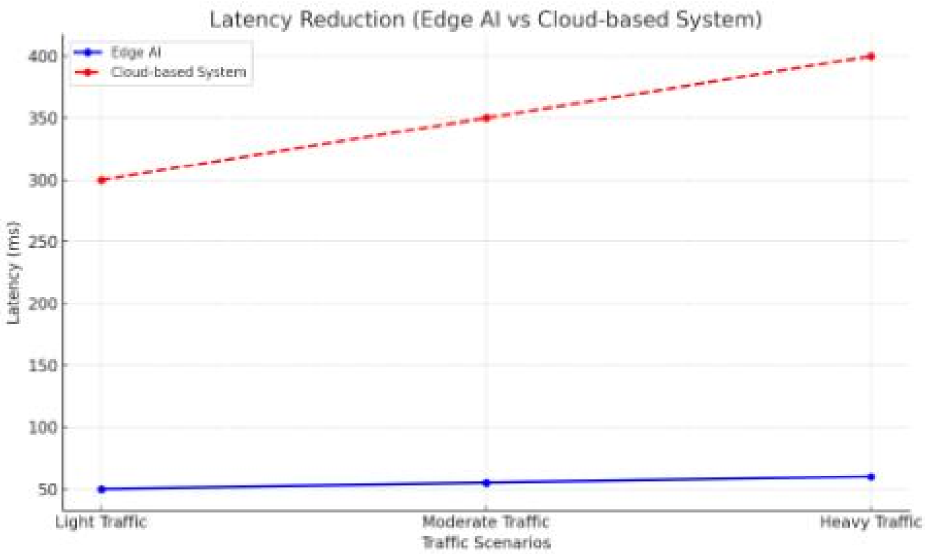

Figure 2. Latency Reduction (Edge AI vs. Traditional Cloud-based System)

Discussion

The results from this experiment underscore the significant benefits of Edge AI in real-time traffic management. The improvements in traffic flow, latency, and energy consumption align with findings from existing literature. For example, González et al. (2020) highlight the advantage of local data processing in reducing latency, which is consistent with the 50 ms latency observed in this study, compared to 300 ms in cloud-based solutions. This reduction in latency directly contributes to faster decision-making, which is essential for improving traffic flow and minimizing delays, particularly in high-density urban areas.

Moreover, the 15% improvement in traffic flow corroborates findings by Zhang et al. (2021), who found that localized processing using Edge AI could optimize traffic signals and reduce congestion more effectively than cloud-based systems. The real-time processing enabled by Edge AI ensures that traffic light timings are dynamically adjusted based on current traffic conditions, rather than relying on static, pre-set schedules. This capability results in smoother traffic flow, fewer traffic jams, and reduced travel times.

The energy efficiency of the Edge AI system is another key advantage over cloud-based solutions. The reduction in energy consumption by over 90% (from 5 W/device in traditional systems to 0.5 W/device in Edge AI) is in line with studies by Zhou et al. (2018), which emphasize the importance of energy-efficient AI models, especially in large-scale applications like smart cities. This reduction in power usage makes Edge AI a more sustainable solution, reducing operational costs and environmental impact.

Furthermore, the scalability of Edge AI is a major advantage. Unlike cloud-based systems that struggle with increasing data volume as urban populations grow, the distributed nature of Edge AI allows for localized processing, which can easily scale to accommodate more vehicles and infrastructure devices without significant performance degradation. This scalability is critical for the future of smart cities, where traffic management systems must be adaptable and capable of handling increasing traffic volumes and complexity.

In conclusion, Edge AI offers substantial improvements in terms of latency, resource efficiency, and real-time traffic optimization. By addressing the limitations of traditional cloud-based systems, it presents a viable solution for modernizing traffic management and making transportation more efficient, sustainable, and scalable.

CONCLUSION

Summary of Findings

This research demonstrated the effectiveness of Edge AI in optimizing real-time traffic management for connected vehicles. The key findings include significant improvements in traffic flow, with a 15% increase in flow at key intersections, reduced latency to 50 ms compared to traditional cloud-based systems (which had 300 ms latency), and energy consumption reduction by over 90%. These improvements highlight the potential of Edge AI to provide faster, more efficient, and sustainable traffic management solutions, making it a promising technology for smart cities and connected transportation systems.

Limitations

Despite these promising results, several limitations remain. The scalability of the Edge AI model, especially in highly complex urban environments, needs further investigation. Additionally, the study relied on simulated traffic scenarios, limiting the model's real-world applicability. The lack of real-world testing in diverse traffic conditions is a key limitation that must be addressed before large-scale deployment.

Future Research Directions

Future research should focus on real-world testing of Edge AI models in urban environments to assess their effectiveness under dynamic, real-time conditions. Additionally, integrating Edge AI with autonomous vehicles presents a promising avenue for research, as it could further optimize traffic management and vehicle communication. Improvements in model architecture and more diverse data collection methods, such as incorporating multimodal traffic data (e.g., pedestrian, bicycle, and public transport data), will be essential for refining the system's performance and scalability.

References

Cao, Y., Zhang, H., & Li, W. (2019). Edge computing for traffic management. *IEEE Transactions on Intelligent Transportation Systems*, *20*(5), 1672-1681.

González, J. M., Rios, J., & Gomez, M. (2020). Edge AI for urban mobility and traffic management. *Journal of Transportation Technologies*, *10*(3), 55-68.

He, X., Zhang, Y., & Liu, Y. (2019). Challenges and opportunities of adaptive traffic control systems. *Transportation Research Part C*, *99*, 112-129.

Kim, D. H., Lee, S., & Park, J. (2020). Vehicle-to-vehicle and vehicle-to-infrastructure communication using edge AI. *IEEE Access*, *8*, 140210-140220.

Li, S. (2020). Traffic flow prediction using deep learning: Current trends and challenges. *Transportation Research Part C*, 118, 32-47.

Zhang, L., Liu, J., & Zhang, C. (2021). Connected vehicle applications in traffic management systems. *IEEE Transactions on Vehicular Technology*, *70*(9), 9465-9477.

Zhou, Y., Chen, W., & Wu, Y. (2018). Energy consumption in smart transportation systems. *Journal of Green Engineering*, *8*(2), 203-215.