

Optimizing embedded AI systems for autonomous driving: Challenges and solutions using bayesian networks

Pan Yuan

University of Sydney, Sydney, Australia

2285985550@qq.com

Abstract. Embedded Artificial Intelligence (AI) systems are important components of autonomous vehicles. However, incorporating AI into autonomous vehicles is technically complex, due to the constraints of computation, real-time processing of data, uncertainty handling, and hardware limitations. Bayesian Networks (BNs) are a promising method that allows probabilistic modelling in adaptive learning and environment perception. Here we report on an overview of the application of BNs on autonomous driving, with an emphasis on how BNs can be optimized for embedded system resource constraints, including both computational and energy. Various optimization techniques are discussed, such as model pruning, approximation, acceleration using hardware accelerators, such as Field-Programmable Gate Arrays (FPGA) and Application Specific Integrated Circuit (ASICs), and advanced cooling and power management to ensure AI reliability under high computational load. By reviewing these approaches, we aim to contribute to the development of more robust and green AI systems for autonomous driving.

Keywords: Embedded AI, Autonomous Driving, Bayesian Networks, Computational Constraints, Real-Time Decision Making.

1. Introduction

Autonomous driving is one of the hottest new technological developments on the horizon, and is set to revolutionise modern transportation systems by improving safety, efficiency and accessibility. A key component of this revolution is the use of embedded artificial intelligence (AI) systems within vehicles, which are required to operate within strict constraints and perform complex tasks such as real-time decision-making, scene/environmental recognition and adaptive learning. Bayesian Networks (BNs), in particular, have found application in this context due to their ability to model uncertainty and make probabilistic inferences from incomplete or noisy data. However, the integration of BNs in the resource-constrained low-power environments of embedded systems imposes a number of challenges, particularly given that most BN models are computationally intensive, and their ability to complete their tasks efficiently is highly dependent on the use of optimised algorithms that can run on the constrained hardware platforms as well as novel power and thermal management techniques to maintain the reliability of the system under high computational loads. This paper will deal with the challenges of implementing BNs in embedded AI systems for autonomous driving. First, we will provide an in-depth analysis of the computational and hardware constraints that impact the performance of BNs. Second, we will cover optimisation techniques and hardware accelerators that can be used to enhance performance [1]. Third, we will address the critical problem of power management and heat dissipation involved in

ensuring that the AI systems run in a sustainable manner in the hostile environment of autonomous vehicles. At the end of this analysis, we will conclude on the effectiveness of BNs in powering autonomous driving technology.

2. Adaptive Learning in Embedded AI Systems

2.1. Computational Constraints

The computational constraints of embedded AI systems render complex models such as Bayesian Networks difficult to implement. However, BNs can be adapted to operate within the confines of an embedded system's processing power through methods such as model pruning and approximation. In this context, the BN serves as a model for all the relevant variables involved in driving, such as speed, road conditions and obstacles.

The BN reflects these relationships as conditional probabilities, such as the probability of one event (whether an obstacle exists given the speed and road condition) given other events (the speed and road condition):

$$P(O|S, R) = \frac{P(O) \cdot P(S|O) \cdot P(R|O)}{P(S) \cdot P(R)} \quad (1)$$

where $P(O|S, R)$ is the conditional probability of an obstacle O given the speed S and road conditions R , $P(O)$ is the prior probability of an obstacle, $P(S|O)$ and $P(R|O)$ are the likelihoods of speed and road conditions given the presence of an obstacle, $P(S)$ and $P(R)$ are the marginal probabilities of speed and road conditions, respectively. Leveraging history, the BN updates those probability distributions to contextualise the current environment, and, by doing so, can make a good guess about what actions the AI should take with its constrained computational capability. For example, time-series data from sensors can update BN nodes by updating the conditional probabilities to recalculate the conditional likelihoods of observable events, and direct computational resources to expectations about the current context's most probable outcome [2]. This way, a BN-based AI system will not only be agile, but will be responsive and accurate enough to work within the computational constraints of an embedded environment.

2.2. Real-Time Decision Making

Another crucial aspect is its ability to make real-time decisions that are required for safe and effective operation of autonomous vehicles. This is especially important when the system must learn to cope with uncertainty. Again, a Bayesian Network provides a structured approach to decision-making under uncertainty. In an embedded AI system, the BN updates its beliefs in real time using incoming sensor data, and uses these beliefs to accelerate the decision-making process. As an example, the BN could model the probabilities of the driving conditions that the vehicle could encounter such as slippery roads or sudden obstacles, and update those probabilities in real time as sensor data is received. In this way, the AI can make a quick assessment, and take action based on the best available information. By focusing the computation on those most likely outcomes, it is possible to achieve real-time functionality without requiring extensive computational power [3]. Furthermore, BNs can handle probabilistic relations that are under-specified, or noisy, which can model the real-world driving conditions that the system will likely encounter, where the data from the sensors may be unreliable, or incomplete.

2.3. Handling Uncertainty

Dealing with this uncertainty is difficult in the context of autonomous driving, but a BN is ideally suited to that objective. An adaptive learning BN is able to modulate the uncertainties presented by the driving environment by representing the probabilistic relationships among sensor data, vehicle-control data and weather data. When a new situation presents itself – for example, the appearance of a pedestrian in the road – the BN uses its probabilistic information to quickly update the likelihood of various outcomes, namely colliding with a pedestrian, or successfully avoid them, and subsequently provide the recommendation for the appropriate type of braking in order to achieve the optimal outcome. By adhering to this basic process, the AI is able to continuously update its probability distribution in order

to dynamically characterise new and unexpected scenarios [4]. In addition to the basic probabilistic reasoning used by a BN, Bayesian Networks can also be coupled with a reinforcement learning algorithm to learn from past decisions and subsequently update their probabilistic distributions to optimise future decisions. For autonomous driving, this gap-bridging innovation allows the AI to learn from new information in order to not only deal with past uncertainties, but to avoid future uncertainties as well. Although there are other mathematical models that can be employed to manage the uncertainty of autonomous driving, the Bayesian Network is ideally suited to autonomous driving due to its ability to dynamically characterise its environment and respond with an appropriate action [5]. In Table 1 below, a Bayesian Network is depicted updating its probability estimates in response to new evidence that an autonomous vehicle will strike a pedestrian.

Table 1. Bayesian Network Probability Updates in Autonomous Driving Scenario

Variable	Initial Probability	New Evidence (Pedestrian Steps into Road)	Updated Probability
(P(Collision \	\Speed, Distance))	0.05	Speed = 50 km/h, Distance = 10 m
(P(Evasive\ Maneuver \	\Speed, Distance))	0.20	Speed = 50 km/h, Distance = 10 m
(P(Safe \ Stop \	\Speed, Distance))	0.75	Speed = 50 km/h, Distance = 10 m
(P(Collision \	\ Weather \ Conditions))	0.02	Weather = Rainy
(P(Evasive \ Maneuver \	\ Weather \ Conditions))	0.10	Weather = Rainy
(P(Safe \ Stop \	\ Weather \ Conditions))	0.88	Weather = Rainy

3. Environmental Perception in Embedded AI Systems

3.1. Sensor Fusion

Environmental perception in autonomous vehicles needs to integrate data from multiple sensors, most notably cameras, light detection and ranging (LiDAR), radar and ultrasonic sensors. Bayesian Networks play an important role in sensor fusion because they provide a probabilistic framework for combining data from these disparate sources. In a BN, each sensor's output is modelled as a node, and dependencies between those nodes are represented as edges. The BN then updates the probabilistic belief about environmental states (such as presence of obstacles, or the state of the road surface) based on the combined information from the sensors. This allows the system to resolve conflicts between sensors and decrease uncertainty in the environmental model [6]. For example, if a radar sensor detects an object at a certain distance, but a camera fails to confirm it, the BN can assign a probability for the presence of the object and then use that information to guide the response of the vehicle. This probabilistic approach to sensor fusion means that the AI system is more likely to create a more accurate and reliable environmental model, when individual sensors provide incomplete or conflicting data.

3.2. Object Detection and Classification

Autonomous driving relies on object detection and classification algorithms, which can be enhanced with Bayesian Networks (BNs). In an embedded AI system, object detection can be modelled as a probabilistic process, where the likelihood of an object being detected (eg, whether it is a pedestrian, vehicle or traffic sign) is linked to sensor data and the context (ie, the current conditions of the environment). For example, the BN can use past experience about the typical appearance and behaviour of pedestrians in different driving conditions to improve the detection of pedestrians. In this way, the BN can update its probabilities in real time based on changing conditions, such as varying light or weather conditions, and maintain high detection accuracy. The BN can also be used to classify objects based on sensor data and contextual information such as the location or movement pattern of the object [7]. With this probabilistic classification approach, the AI system can make better decisions about how

to respond to different types of objects, minimising false positives or negatives and maximising safety [8].

3.3. Environmental Mapping

Accurate environmental mapping is critical for autonomous driving, and Bayesian Networks provide a robust framework for this task. In a BN-based mapping system, the environment is represented as a probabilistic model, where different features—such as roads, buildings, and obstacles—are nodes connected by edges that represent spatial relationships. As the vehicle moves, the BN updates its map based on sensor data, adjusting the probabilities of different features and their locations. This real-time updating capability allows the AI system to maintain an accurate and up-to-date map, even in dynamic environments where conditions can change rapidly. For example, if the vehicle detects a new obstacle that was not present in the initial map, the BN can update the map to include this obstacle and adjust the probabilities of other features accordingly. This probabilistic approach to environmental mapping ensures that the AI system can navigate complex environments with greater confidence and accuracy, supporting safer and more reliable autonomous driving [9]. Figure 1 represents the initial and updated probabilities of different environmental features, as adjusted by a Bayesian Network during real-time environmental mapping in an autonomous driving system.

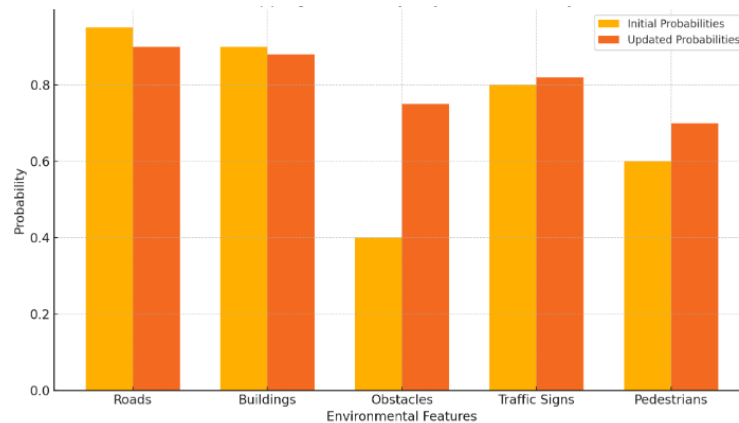


Figure 1. Environmental Mapping: Probability Adjustments in Bayesian Network

4. Integration of Embedded AI with Vehicle Hardware

4.1. Hardware Constraints

Bayesian Networks, with their inherent flexibility and extendibility, pose a challenge in constraint-heavy environments such as the embedded AI of autonomous vehicles. Although the power of BNs is great, inference and updating probabilities needs considerable computational resources, which may require real-time constraints to be met in the embedded environment. Keeping in mind the need for enhancing the performance of BNs and integrating them within embedded AI systems, researchers have developed several variants of BNs along with various methods and algorithms. These optimized BN algorithms intend to overcome the computational constraints of hardware, and implement BNs in the embedded AI system of an autonomous vehicle. The optimizations in these BNs may have various forms such as simplifying the network structure, reducing the nodes and edges, or approximating the model through inference. Another way to implement this is through the design of hardware accelerators like Field Programmable Gate Arrays (FPGAs) and Application-Specific Integrated Circuits (ASICs) to be computationally efficient and fit the physical and computing constraints of the vehicle itself [10]. Figure 2 Overview of Bayesian Network Structure, Learning Process and Implementation Flow in a Constrained Environment.

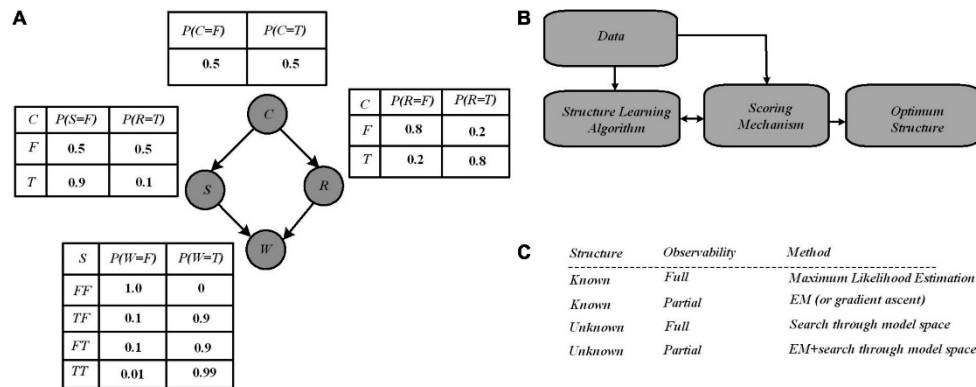


Figure 2. Optimized Bayesian Network Implementation and Learning Process in Embedded AI Systems (Source:Frontiers.com)

4.2. Power Management

In electric vehicles, for instance, battery life is a big concern even in the most mundane driving conditions. Such computationally expensive Bayesian Networks can be made more energy-efficient by using traditional optimisation techniques to minimise the overall power consumption. One line of research centres around dynamic voltage and frequency scaling (DVFS), which enters the picture when the hardware that runs the BN algorithms has multiple operating modes. For instance, in the case of an electric vehicle, this means reducing the frequency of BN updates when the vehicle is cruising on the highway with little traffic, as a way to reduce power consumption, then increasing it when it's navigating a busy intersection, to ensure it can update quickly enough to track the changing situation on the road. Another line of research suggests that Bayesian Networks could be more energy-efficient on embedded chips by using energy-efficient BN models, such as ones based on binary decision diagrams or sparse representations. Power management is critical to ensure that embedded AI systems can operate efficiently and sustainably in the energy-constrained operating environment of an autonomous vehicle.

4.3. Heat Dissipation

Heat dissipation is a significant challenge in the integration of embedded AI systems, particularly when running computationally intensive models like Bayesian Networks. The high processing demands of BNs generate substantial heat, which must be managed effectively to prevent overheating and ensure the reliability of the system. To address this, advanced cooling techniques, such as liquid cooling and heat pipes, are being explored for use in automotive environments. These cooling solutions can provide efficient heat dissipation in compact spaces, ensuring that the AI system remains within safe operating temperatures even during high computational loads. Additionally, thermally optimized hardware, such as AI accelerators with integrated heat management features, can help reduce the thermal load on the system. For example, by distributing the computational workload across multiple processing units with individual cooling mechanisms, the system can prevent localized hotspots and improve overall thermal efficiency. Combined with software-based thermal management strategies—such as workload balancing and thermal-aware scheduling—these hardware solutions ensure that embedded AI systems can operate reliably and safely in the demanding environment of autonomous driving.

5. Conclusion

Those embedded AI systems that adopt BNs for autonomous driving would have a superior capability to reason probabilistically in real time under uncertain and dynamic driving conditions. The aforementioned key challenges would certainly complicate the design space for adopting BNs in embedded AI systems, including the computational constraints imposed by the microprocessor units or ASICs, the hardware constraints that limit the number of BN nodes and interconnections, and the hardware constraints that require efficient power and thermal dissipation for system reliability.

Nonetheless, the probabilistic reasoning and adaptive learning nature of BNs would provide a robust means to address these challenges. Above all, the optimisation techniques that include model pruning, approximation methods and hardware accelerators would ensure the feasibility of embedding BNs in the extremely tight embedded system constraints. These innovative cooling and power management strategies would be another key part of assuring the reliability and efficiency of the AI systems operating in high-stress environments. The advent of autonomous driving would certainly require the adoption of intelligent BNs that can manage the uncertainty of dynamic driving conditions and reduce risk for safe driving. The design challenges and optimisation techniques discussed in this paper contribute to realising robust embedded AI systems for autonomous driving in the future.

References

- [1] Sonko, Sedat, et al. "A comprehensive review of embedded systems in autonomous vehicles: Trends, challenges, and future directions." *World Journal of Advanced Research and Reviews* 21.1 (2024): 2009-2020.
- [2] Harrison, Klasan, Roopak Ingole, and Srinivas Naveen Reddy Dolu Surabhi. "Enhancing Autonomous Driving: Evaluations Of AI And ML Algorithms." *Educational Administration: Theory and Practice* 30.6 (2024): 4117-4126.
- [3] Gutiérrez-Zaballa, Jon, et al. "On-chip hyperspectral image segmentation with fully convolutional networks for scene understanding in autonomous driving." *Journal of Systems Architecture* 139 (2023): 102878.
- [4] Biglari, Amin, and Wei Tang. "A review of embedded machine learning based on hardware, application, and sensing scheme." *Sensors* 23.4 (2023): 2131.
- [5] Thieu, Gia Bao, et al. "ZuSE Ki-Avf: application-specific AI processor for intelligent sensor signal processing in autonomous driving." *2023 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2023.
- [6] Conrad, Christopher, Saba Al-Rubaye, and Antonios Tsourdos. "Intelligent embedded systems platform for vehicular cyber-physical systems." *Electronics* 12.13 (2023): 2908.
- [7] Naya-Varela, Martin, et al. "Robobo smartcity: An autonomous driving model for computational intelligence learning through educational robotics." *IEEE Transactions on Learning Technologies* 16.4 (2023): 543-559.
- [8] Cummings, M. L. "Revising human-systems engineering principles for embedded AI applications." *Frontiers in Neuroergonomics* 4 (2023): 1102165.
- [9] Hasanujjaman, Muhammad, Mostafa Zaman Chowdhury, and Yeong Min Jang. "Sensor fusion in autonomous vehicle with traffic surveillance camera system: detection, localization, and AI networking." *Sensors* 23.6 (2023): 3335.
- [10] Miraliev, Shokhrukh, et al. "Real-time memory efficient multitask learning model for autonomous driving." *IEEE Transactions on Intelligent Vehicles* 9.1 (2023): 247-258.