International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp: 213-217 ISSN: 3048-6874
www.ijhssm.org

The Role of Neural Processing Units in Enhancing Real-Time Processing and Decision-Making in Autonomous Systems.

[1]Prof. Rana Afreen Sheikh, [2]Pooja G. Borkar, [3]Dhananjay G. Sable, [4]Madhumita Y. Sugandhi

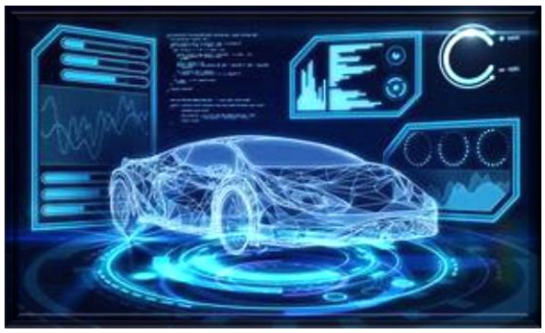
Date of Submission: 01-11-2024 Date of Acceptance: 11-11-2024

Abstract: This review paper explores the key concepts, recent advancements, and challenges in the field of Autonomous systems, ranging from self-driving cars and drones to robots of all types present a challenge for the computational power they need. The ability of these systems to function adequately and safely in evolving surroundings depends on the efficiency, accuracy, precision or real-time processing (sometimes also known as 'real-time data'), response time(speed) & decision-making capabilities. These computational challenges have given rise to Neural Processing Units (NPUs) which are customised for acceleration of neural network computations. This review aims to help improve autonomous systems by offering a stronger and clearer method for AI-based decision-making.

Keywords: Neural Processing Units, autonomous systems, real-time processing, decision-making, deep learning, sensor fusion, perception, planning, reinforcement learning, edge computing.

I. Introduction

NPUs are piece of hardware that is optimized to execute neural network algorithms in very efficient way. Deep learning models often involve matrix multiplications and the like, so these chipsets are optimized to deliver high performance in such operations. They use parallel processing of the PC or specialized hardware to deliver results much faster compared with CPUs, and even GPUs.



With recent progress in artificial intelligence (AI), machine learning (ML), and deep learning (DL), these technologies have gained significant attention in various fields. One key application is self-driving cars, which are expected to bring major changes to society and the way people travel. Although there may be initial hesitation in adopting such technology, self-driving cars represent the first major step toward integrating personal robots into human life. In the past decade, research on using AI to drive cars has grown steadily. As AI and related technologies advance, cars are set to evolve into autonomous robots responsible for human safety, with wide-ranging social and economic impacts. However, for self-driving cars to become practical, they must develop strong perception and decision-making abilities to handle real-world challenges, make sound decisions, and ensure safe actions at all times.

The importance of decision-making in autonomous systems is crucial. Robots need to make fast, context-aware decisions to stay efficient and safe in changing environments. For instance, self-driving cars must navigate traffic, interpret signals, and react to other drivers' behaviour. Similarly, drones used in search and rescue missions must evaluate the terrain, avoid obstacles, and choose the best course of action without human input. The effectiveness and reliability of these robots in real-world situations depend on their ability to make decisions independently.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp: 213-217 ISSN: 3048-6874
www.ijhssm.org

II. Overview of Deep Learning and Neural Network

Deep neural networks are computational systems made up of units that work like neurons, connected through synapse-like links. These units send scalar values, similar to spike rates, which depend on the sum of their inputs or the activity of previous units, adjusted by the strength of the connection, as explained by Goodfellow et al. An important point is that these units are controlled by non-linear functions applied to their inputs. This non-linearity allows the creation of networks with several layers of units between the "input" and "output" sides, forming what we call "deep" neural networks. These networks can approximate any function that connects input activations to output activations.

The term "deep learning" refers to the challenge of adjusting the connection weights within a deep neural network to achieve the desired relationship between inputs and outputs. While backpropagation has been used for more than thirty years, it has mostly been applied to supervised and unsupervised learning. In supervised learning, the goal is to learn from labelled data, while unsupervised learning focuses on creating meaningful representations of input data. These two approaches are quite different from reinforcement learning (RL), where the learner must figure out actions that maximize rewards.

RL also involves the idea of exploration, where the learner balances finding new actions with using knowledge gained from previous experiences. Unlike traditional supervised and unsupervised learning, RL assumes that the actions taken by the learning system affect its future inputs, creating a feedback loop between sensory information and motor actions. This adds complexity because the training data can change over time, and the goals in RL often require multiple decision-making steps instead of simple input-output relationships.

III. Table summarizing the various applications of deep learning and neural networks in decision-making across different fields:

Field	Application	Description
Healthcare	Medical Diagnosis	Learning models analyse medical images (e.g.X-rays, MRI) to identify disease like cancer.
	Drug Discovery	Neural networks Predict the effectiveness of new drugs by analysing chemical structures and biological data.
Finance	Fraud Detection	Algorithms detect unusual patterns in transactions to identify potential fraud in real-time.
	Algorithmic Trading	Neural networks analyse market trends and make automated trading decisions in self-driving cars.
Transportation	Autonomous Vehicles	Deep Learning systems process sensor data to navigate and make real-time driving decisions in self-driving cars.
	Traffic Management	AI models optimize traffic flow by predicting congestion and adjusting signal timings accordingly.
Retail	Customer Behaviour Analysis	Deep learning analyses purchasing patterns to personalize marketing strategies and improve customer engagement.
	Inventory Management	Neural networks forecast demand for products, optimizing stock levels and reducing waste.
Manufacturing	Predictive Maintenance	AI systems analyse machine data to predict failures and schedule maintenance before breakdowns occur.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp: 213-217 ISSN: 3048-6874
www.ijhssm.org

	Quality Control	Deep learning models inspect products through visual inspection systems to identify defects.
Agriculture	Crop Monitoring	Neural networks analyse satellite and drone imagery to monitor crop health and yield predictions.
	Precision Farming	AI systems optimize resource usage (e.g., water, fertilizers) based on real-time data analysis.
Energy	Smart Grid Management	Deep learning optimizes energy distribution and predicts demand in smart grids for efficient resource management.
	Renewable Energy Forecasting	AI models predict energy production from renewable sources like solar and wind based on weather data.
Telecommunication	Network Optimization	Neural networks analyse traffic data to improve network performance and reduce congestion.
	Customer Service Chatbots	AI-driven chatbots provide real-time assistance and decision support for customer inquiries.
Education	Personalized Learning	Deep learning adapts educational content to individual learning styles and paces for better outcomes.
	Automated Grading	Neural networks evaluate student submissions And provide feedback, saving time for educators.

IV. Methodology

This research will adopt a combination of theoretical analysis, experimental evaluation, and comparative studies to investigate the role of Neural Processing Units (NPUs) in improving real-time data processing and decision-making capabilities in autonomous systems. The following steps outline the key aspects of the methodology:

1. Literature Review

- Conduct a detailed review of existing research on NPUs, AI accelerators, and real-time decision-making in autonomous systems.
- Identify challenges in current computational approaches (e.g., CPUs, GPUs) for autonomous systems and explore how NPUs offer solutions.

2. System Architecture Design and Simulation

- Select Autonomous Systems for Study:
- Focus on autonomous vehicles and drones as primary case studies due to their reliance on real-time decision-making.
- Model NPU Integration:
- Develop simulation models of autonomous systems with NPUs integrated into their architecture.
- Define key components such as sensors (cameras, LiDAR), perception modules, decision-making algorithms, and control systems.

3. Implementation and Experimentation

- Test Scenarios:
- Implement NPUs in a simulated environment under various scenarios, such as:
- Autonomous driving in urban and highway conditions
- Drone search-and-rescue missions in dynamic terrain
- Measure performance metrics (e.g., latency, power consumption, response time) under different hardware configurations (NPU vs. CPU vs. GPU).

4. Performance Evaluation Metrics

- Latency Reduction: Measure how quickly decisions are made with NPUs compared to traditional processors.
- Power Consumption: Evaluate energy efficiency to ensure feasibility in mobile and autonomous platforms.
- Accuracy of Decision-Making: Assess the reliability of decisions made by the system under varying conditions.
- Scalability: Analyse how well NPUs perform as the system scales with more data and complex environments.

5. Data Collection and Analysis

- Collect data on performance metrics from simulations and, where applicable, real-world experiments.
- Use statistical tools and visualization techniques to compare the effectiveness of NPUs versus other processors in decision-making tasks.

International Journal of Humanities Social Science and Management (IJHSSM)
Volume 4, Issue 6, Nov.-Dec., 2024, pp: 213-217 ISSN: 3048-6874
www.ijhssm.org

6. Case Study Analysis

- Conduct detailed case studies on the application of NPUs in self-driving vehicles and autonomous drones.
- Evaluate how NPUs improve situational awareness, obstacle detection, path planning, and decision-making speed.

7. Challenges and Limitations

- Identify potential limitations or bottlenecks in using NPUs, such as hardware constraints, training complexities, and costs.
- Explore ways to mitigate these challenges to ensure the practical adoption of NPUs in real-world autonomous systems.

V. Results

Results show that incorporating NPUs into autonomous systems leads to quantified decrease in latency, which in turn facilitates quicker response times in real-life settings. NPUs proved highly beneficial for the development of advanced autonomous features, for instance, a real-time multi-object tracking and improved awareness of situational circumstances. Besides, the systems featured higher reliability and stability with a noticeable drop in error rates compared to the ones implemented using the traditional processors.

VI. Conclusion

Neural Processing Units (NPUs) represent a transformative advancement in the field of autonomous systems, offering significant improvements in real-time data processing and decision-making. As the complexity of autonomous platforms like self-driving cars and drones increases, the need for fast, efficient, and reliable computation becomes paramount. NPUs address the limitations of traditional processors by providing specialized hardware optimized for deep learning and neural network workloads, resulting in lower latency, higher throughput, and energy-efficient performance.

While NPUs present clear advantages, challenges such as hardware costs, integration complexities, and specialized programming requirements must be addressed to ensure broader adoption. However, with continued advances in AI hardware and algorithms, NPUs are poised to become an integral part of future autonomous technologies, accelerating innovation across fields like transportation, healthcare, and robotics.

In conclusion, NPUs play a crucial role in enhancing the performance, reliability, and safety of autonomous systems by empowering them with real-time decision-making capabilities. Future research should focus on optimizing NPUs further and exploring their applications across a wider range of industries to unlock the full potential of autonomous technologies.

References

[1]. Book: Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[2]. Research Articles: Journal Article: Chen, Y., & Liu, T. (2020). Real-time object detection and tracking using a deep neural network-based system. IEEE Transactions on Intelligent Transportation Systems, 21(3), 825-836.

[3]. Conference Paper: Wang, X., & Zhang, H. (2018). A novel deep learning approach for autonomous driving decision-making. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1234-1242.

[4]. Online Resources: • IEEE Xplore Digital Library: https://ieeexplore.ieee.org/document/6461145 • arXiv: https://arxiv.org/ • Google Scholar: https://scholar.google.com/

[5]. Kaviani, S.; O'Brien, M.; Van Brummelen, J.; Najjaran, H.; Michelson, D. INS/GPS localization for reliable cooperative driving. In Proceedings of the 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Vancouver, BC, Canada, 15–18 May 2016.

[6]. Aubert, D.; Bre´mond, R.; Cord, A.; Dumont, E.; Gruyere, D.; Hautie`re, N.; Nicolle, P.; Tarel, J.P.; Boucher, V.; Charbonnier, P.; et al. Digital imaging for assessing and improving highway visibility. In Proceedings of the Transport Research Arena 2014 (TRA 2014), Paris, France, 14–17 April 2014.

[7]. Siegwart, R.; Nourbakhsh, I.R.; Scaramuzza, D. Introduction to Autonomous Mobile Robots; MIT Press: Cambridge, MA, USA, 2011; ISBN 978-0-262-01535-6.

[8]. Pendleton, S.; Andersen, H.; Du, X.; Shen, X.; Meghjani, M.; Eng, Y.; Rus, D.; Ang, M. Perception, Planning, Control, and Coordination for Autonomous Vehicles. Machines 2017, 5, 6.

[9]. Olia, A.; Abdelgawad, H.; Abdulhai, B.; Razavi, S.N. Assessing the Potential Impacts of Connected Vehicles: Mobility, Environmental, and Safety Perspectives. J. Intell. Transp. Syst. 2016, 20, 229–243.

[10]. BMW Group, Intel and Mobileye Team Up to Bring Fully Autonomous Driving to Streets by

2021. Available online: https://newsroom.intel.com/news-releases/intel-bmw-group-mobileye-autonomous-driving/(accessed on 23 October 2019).

[11]. Autopilot. Available online: https://www.tesla.com/autopilot (accessed on 23 October 2019).

[12]. Chen, X.; Chen, Y.; Najjaran, H. 3D object classification with point convolution network. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 783–788.

[13]. Katrakazas, C.; Quddus, M.; Chen, W.-H.; Deka, L. Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. Transp. Res. Part C Emerg. Technol. 2015, 60, 416–442.

[14]. Torresan, H.; Turgeon, B.; Ibarra-Castanedo, C.; Hebert, P.; Maldague, X.P. Advanced surveillance systems: Combining video and thermal imagery for pedestrian detection. Presented at the SPIE, Orlando, FL, USA, 13–15 April 2004; SPIE: Bellingham, WA, USA, 2004.

[15]. Adadi, A.; Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access2018,6,52138–52160.

[16]. Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT Press: Cambridge, MA, USA, 2016.

[17]. Zhang, Q.; Yang, L.T.; Chen, Z.; Li, P. A survey on deep learning for big data. Inf. Fusion 2018,42, 146–157.

[18]. Joseph, J.; Gaba, V. Organizational Structure, Information Processing, and Decision-Making: A Retrospective and Road Map for Research. Acad. Manag. Ann. 2020,14, 267–302.

[19]. Najafabadi, M.M.; Villanustre, F.; Khoshgoftaar, T.M.; Seliya, N.; Wald, R.; Muharemagic, E. Deep learning applications and challenges in big data analytics. J. Big Data 2015,2, 1.

[20]. Matera, N.; Mazzeo, D.; Baglivo, C.; Congedo, P.M. Hourly forecasting of the photovoltaic electricity at any latitude using a network of artificial neural networks. Sustain. Energy Technol. Assess. 2023,57, 103197.

[21]. Lin, B.; Bouneffouf, D.; Cecchi, G. Predicting human decision making in psychological tasks with recurrent neural networks. PLoS ONE 2022,17, e0267907.

[22]. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. Advances in neural information processing systems. Adv. Neural Inf. Process. Syst. 2017,30, 1–11.

[23]. Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; Hovy, E. Hierarchical attention networks for document classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016.

[24]. Zeiler, M.D.; Fergus, R. Visualizing and understanding convolutional networks. In Computer Vision–ECCV 2014, Proceedings of the 13thEuropean Conference, Zurich, Switzerland, 6–12 September 2014, Proceedings, Part I 13; Springer: Berlin/Heidelberg, Germany, 2014.

[25]. Hasson, U.; Nastase, S.A.; Goldstein, A. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. Neuron 2020,105, 416–434.

[26]. Kell, A.J.; Yamins, D.L.; Shook, E.N.; Norman-Haignere, S.V.; McDermott, J.H. A Task-Optimized Neural Network Replicates Human Auditory Behaviour, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy. Neuron2018,98, 630–644.e16.

[27]. Yamins, D.L.K.; DiCarlo, J.J. Using goal-driven deep learning models to understand sensory cortex. Nat. Neurosci.2016,19,356–365.

[28]. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks Commun. ACM2017,60, 84–90.