2026 IEEE/SICE International Symposium on System Integration (SII)
January 11-14, 2026. Cancún, México

Offloading Reservoir Computing-Based Hand-Waving Action Recognition to FPGAs for Service Robots

Sojiro Shimada[1], Akinobu Mizutani[1], Kanta Yoshioka[1,2], Kosei Isomoto[1], and Hakaru Tamukoh[1,3]

***Abstract*— We propose a system that offloads hand-waving action recognition processing in service robots to a field-programmable gate array (FPGA) accelerator. The accelerator implements a LUTNet-based reservoir computing (LUTNet-RC) model. Conventionally, such processing was performed by laptops connected to robots. Nevertheless, the proposed system offloads some of these task processing to an FPGA. This improves the overall efficiency of the robot and optimizes the use of computational resources. The experimental results show that in hand-waving action recognition, the LUTNet-RC on FPGA achieves higher accuracy and precision than the conventional echo state network (ESN) model, while significantly reducing the power consumption. In addition, the system can be expanded to large-scale networks, which are challenging to implement on laptops.**

I. INTRODUCTION

Demographic shifts in developed countries, including declining birth rates and aging populations, cause to labor force shortages. Under such conditions, the service robot market has expanded and is expected to continue to grow [1]. Among the various types of service robots, domestic service robots specializing in tasks, such as tidying up rooms and delivering objects to humans, are being actively researched and developed [2]. Voice interactions are commonly used in commercial service robots and smart speakers, where wake-up words are used to trigger commands [3]. However, voice input often fails to function reliably in noisy environments or in situations where speaking is difficult. As alternative non-verbal approaches, techniques such as hand-waving detection have been used as wake-up methods under such circumstances [4]. Several technical problems must be solved to achieve advanced processing, such as hand-waving action recognition on robots. In contrast, methods using deep neural networks (DNNs) achieve high recognition accuracy; however, they require substantial computing resources and large memory capacity, making it difficult to install them on robots with limited resources [5]. To address this problem, a lightweight calculation method using echo state network (ESN), a type of reservoir computing (RC), has been proposed [6], [7]. These lightweight methods are executed on the laptop central processing unit (CPU), and thus, problems related to processing cost and resource competition persist. Robots perform multiple tasks, such as self-localization and object detection, in parallel, and the CPU load and memory usage caused by recognition processing may affect the overall real-time performance. Furthermore, constant CPU processing increases the power consumption, which limits the operating time of battery-powered robots.

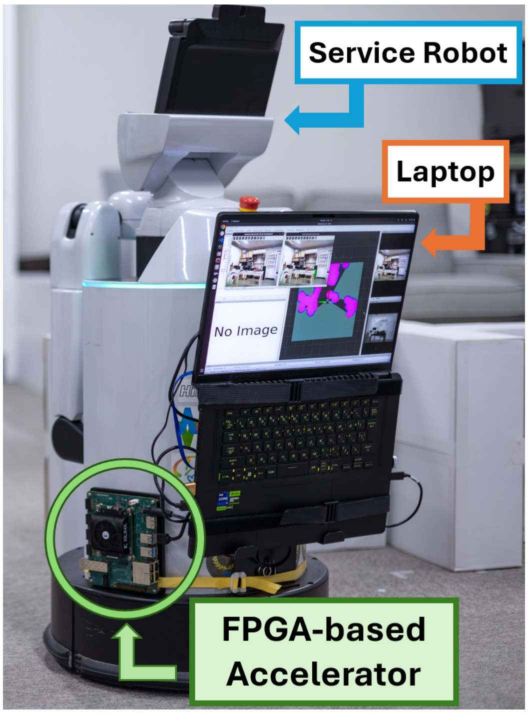

Fig. 1. Service Robot with Laptop and FPGA-based Accelerator

To solve these problems, we propose a system that offloads hand-waving action recognition processing to a dedicated hardware accelerator. The hardware-oriented RC [8] was implemented in field-programmable gate arrays (FPGAs), enabling the offloading of the processing tasks that were previously performed on a laptop CPU. As shown in Fig. 1, the proposed system was implemented by integrating an FPGA-based accelerator into a service robot equipped with a laptop. The contributions of this study are as follows.

This research is based on results obtained from a project, JPNP16007, commissioned by the New Energy and Industrial Technology Development Organization (NEDO). This work received support from JSPS KAKENHI Grant Numbers 23H03468 and 24KJ1820, as well as from JST ALCA-Next Grant Number JPMJAN23F3.

[1] All authors are with Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu, 808-0196, Japan `shimada.sojiro230@mail.kyutech.jp`, `mizutani.akinobu515@mail.kyutech.jp`, `yoshioka.kanta986@mail.kyutech.jp`, `isomoto.kosei778@mail.kyutech.jp`

[2] Kanta Yoshioka is also with Japan Society for the Promotion of Science, 5-3-1 Kojimachi, Chiyoda, Tokyo, 102-0083, Japan `yoshioka.kanta986@mail.kyutech.jp`

[3] Hakaru Tamukoh is also with Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu, Kitakyushu 808-0196, Japan `tamukoh@brain.kyutech.ac.jp`

978-1-6654-5784-2/26/$31.00 ©2026 IEEE

- We proposed a system that offloads the processing of RC-based hand-waving action recognition to an FPGA-based accelerator for service robots.
- We reduced power consumption and memory usage by performing the task processing on the proposed FPGA-based accelerator instead of using the laptop CPU.
- We demonstrated that, in the hand-waving action recognition task, the hardware-oriented RC on an FPGA in the proposed system achieved higher recognition accuracy and precision than the ESN.
- We confirmed that frequently used tasks, such as hand-waving recognition, can be offloaded to an FPGA accelerator, reducing the CPU load and dependence on high-performance laptops or cloud resources.

II. RELATED WORKS

A. Field-Programmable Gate Array

FPGAs are reconfigurable semiconductor devices that can be programmed to configure their internal circuit structures. They consist of programmable logic blocks and reconfigurable interconnects, enabling flexible implementation of various digital circuits. FPGAs mainly comprise of look-up tables (LUTs) and flip-flops (FFs). LUTs store arbitrary truth tables in small-scale memory and generate outputs based on inputs; thus, they are used in the implementation of combinational circuits. In contrast, FFs are memory elements that store a single bit of data synchronized with a clock signal and are used in the construction of sequential circuits [5].

FPGAs can achieve low power consumption and high-performance processing by optimizing data flow for specific applications. They also utilize high parallelism, unlike general-purpose processors such as CPUs and graphics processing units (GPUs) [9]. This characteristic enables the implementation of accelerators for real-time data processing in edge computing environments [10].

B. Reservoir Computing

An RC is a type of recurrent neural network (RNN) used to process time-series data. The RC consists of three layers: an input layer with N_i neurons, a reservoir layer with N_r neurons, and an output layer with N_o neurons. In RC learning, the weights W_{ir} from the input layer to the reservoir layer and recurrent weights W_{rr} within the reservoir layer are initialized randomly and fixed. The learning target is limited to the weight W_{ro} from the reservoir layer to the output layer, and W_{ro} is trained using a lightweight algorithm, such as linear regression, which keeps the learning cost low. With this structure, RC features a simple network architecture and low learning cost, while achieving a high accuracy comparable to that of conventional fully-trained RNNs [6]. Based on these characteristics, RC is being investigated for applications in edge computing, and hardware implementation using FPGAs is being promoted [11], [12], [13], [14], [15]. However, the calculation cost of accumulation in reservoir layers remains high. Therefore, it is important to optimizeoptimizing the calculation efficiency of the FPGA implementation [8].

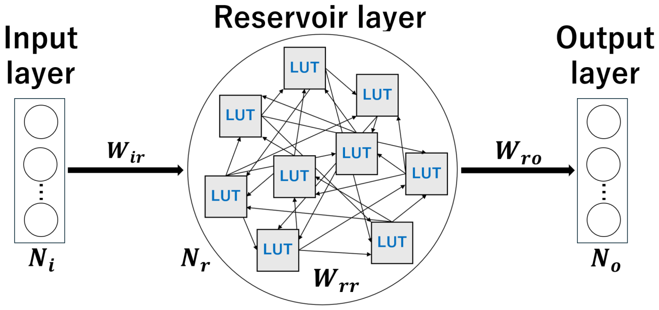

Fig. 2. LUTNet-RC model

One approach to solving this problem in FPGA implementations is LUT networks-based RC (LUTNet-RC) [8]. LUTNet is a quantization method that specializes in FPGA architecture. It defines a neuron model based on the LUTs, which are the fundamental elements of an FPGA, and constructs a network. As shown in Fig. 2, LUTNet-RC is a model that applies LUTNet to the reservoir layer of RC. By constructing the reservoir layer using LUTNet, the calculation cost of multiply-accumulation operations in the reservoir layer is significantly reduced, allowing the computation to be completed within a single clock cycle [8]. Conventionally, implementing the arithmetic circuit of each neuron required hundreds of LUT elements to build adders. In contrast, LUTNet maps the input–output relationship of each neuron directly onto a single LUT element, ideally enabling one neuron to be implemented with one LUT [5].

C. Service Robots

Service robots are expected to efficiently perform various tasks, such as caring for the elderly and assisting with housework. To work smoothly with humans, robots must accurately interpret human intentions. This is especially true for recognizing nonverbal cues, such as gestures. For example, in RoboCup@Home, an international robotics competition focused on service robot research and development, some tasks include elements in which robots respond to user's hand actions [16].

However, it is difficult for service robots to run such recognition functions continuously over long times, because the onboard computer is battery-powered and computing resources are strictly constrained. Service robots require an ability to assess situations to act autonomously, and thus use multiple vision systems that run on a computer with a GPU mounted on the robot [17]. Although a GPU can be used for high-precision recognition processing, this increases power consumption and limits the robot's operating time. Furthermore, the robot's control system must execute multiple tasks in parallel, including self-localization, object detection, and object grasping. In RoboCup@Home, these processes have been reported to be executed in parallel using CPUs and GPUs [17]. In such a situation with limited resources, using the laptop CPU for continuously running processes, such as hand-waving action recognition, may strain existing task processing and reduce the responsiveness of the system as

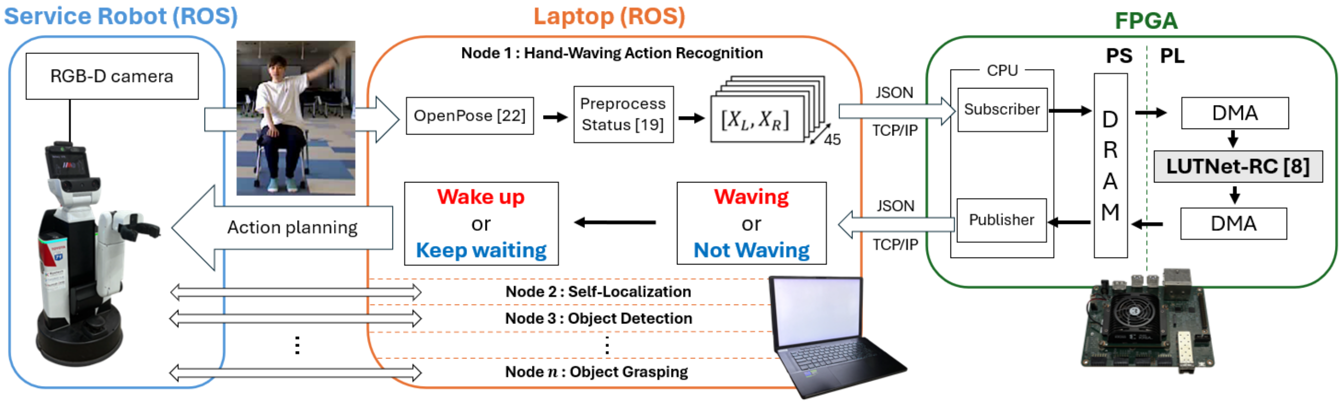

Fig. 3. Proposed System Architecture

a whole. Therefore, offloading specific processes with high calculation loads to specialized hardware accelerators with low power consumption is an effective way to maintain performance and achieve long operating times in service robots.

III. Proposed System

A. Overall System Architecture

As shown in Fig. 3, the proposed system consists of three main components: a service robot and a laptop, both running the robot operating system (ROS) [18], and an FPGA-based accelerator. The service robot is equipped with an RGB-D camera that captures video sequences of the scene. These images are processed on the laptop to extract skeletal information and convert it into time-series data (X_L, X_R) representing left and right hand movements [19]. The laptop executes various processing tasks in parallel, and the hand-waving action recognition task is offloaded to the FPGA accelerator. The laptop transfers the formatted data to the FPGA via a Ethernet cable using JavaScript Object Notation (JSON) over Transmission Control Protocol/Internet Protocol (TCP/IP). The FPGA performs recognition processing and returns the results to the laptop, which sends them back to the service robot. This system can be applied to highly optimized ROS-FPGA communication systems [20], enabling efficient data exchange between ROS on the laptop and the FPGA via a TCP/IP over a Ethernet, simply by replacing the internal program of the FPGA.

The RC accelerator circuit on the FPGA is implemented on the PYNQ [21] platform provided by AMD Xilinx. Therefore, the RC circuit inside programmable logic (PL) can be controlled from the CPU on the processing system (PS) using Python. Robot systems are often developed using Python because of their high compatibility with artificial intelligence (AI) and robot operating systems. In contrast, circuits on FPGAs are often designed using low-level languages, such as C. However, this system simplifies the integration of FPGA accelerators into robot systems and reduces development costs by standardizing Python.

B. Preprocessing on Laptop

As in conventional methods [19], the preprocessing stage of this system uses video sequences captured by a camera as input to extract skeletal information and generate time-series data for recognition. To extract the skeletal information, we used OpenPose [22]. OpenPose can detect the positions of key points on the body, such as the wrist and shoulder, in real time from each frame of the video sequences. The obtained coordinate values are normalized to reduce the effect of differences in person position and size on the recognition accuracy. Specifically, the coordinates of the nose are used as reference points, and the distance between the left and right shoulders is used as the scale to normalize the wrist coordinates [19]. This process generates two time-series $[X_L, X_R]$ for the left and right hands, respectively. These time-series are used as inputs for the FPGA accelerator.

C. LUTNet-RC Accelerator on FPGAs

The proposed LUTNet-RC accelerator was implemented as a specialized hardware circuit in the PL area of a system-on-a-chip (SoC). This accelerator consists of two blocks: an LUTNet-based reservoir layer and an output layer. Data exchange with the PS side is performed by a direct memory access (DMA) controller using shared dynamic random access memory (DRAM). Using LUTNet as a reservoir layer preempts the requirement of, large-scale multiply-accumulation circuits, making implementation possible with extremely small circuit resources [8], [23].

IV. Experiments

A. Circuit Size

In this experiment, we evaluated the circuit size while changing the number of neurons (N_r) in the reservoir layer of the proposed LUTNet-RC accelerator. Specifically, we created LUTNet-RC intellectual property (IP) cores with N_r set to 128, 256, 512, 1024, 2048, and 4096. Each IP core was incorporated into the block design of the AMD Xilinx Vivado v2023.2, and implemented on the target board KR260 (device: Zynq UltraScale+ MPSoC EV (XCK26)). After implementation, we checked circuit size of the proposed

TABLE I

Circuit Size of Proposed LUTNet-RC system

N_r	LUT (%)	FF (%)	BRAM (%)
4096	14,510 (12.39)	25,102 (10.72)	2 (1.39)
55000†	116,230 (99.24)	226,416 (96.66)	–

† Estimated based on linear increase with changes in circuit scale N_r

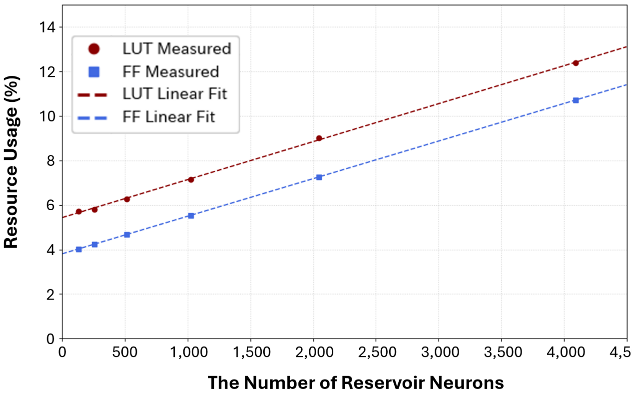

Fig. 4. Estimation of FPGA Resource Usage

TABLE II

Accuracy and Precision

N_r	LUTNet-RC		ESN [19]	
	Accuracy	Precision	Accuracy	Precision
128	**0.73611**	**0.79487**	0.63194	0.61314
256	**0.75000**	**0.80769**	0.67361	0.65041
512	**0.75694**	**0.86567**	0.68056	0.68269
1024	**0.77778**	**0.88235**	0.67361	0.65546
2048	**0.78472**	**0.83544**	0.72222	0.68644
4096	**0.80556**	**0.85000**	0.68750	0.65354

LUTNet-RC accelerator. The results for $N_r = 4096$ are shown in Table I.

As the circuit scale was primarily dominated by the usage of LUTs and FFs, we analyzed the changes in usage with increasing N_r. Consequently, it was confirmed that the use of both LUTs and FFs increased approximately linearly. As shown in Fig. 4, based on this linear relationship, we estimated that the maximum number of neurons that can be implemented on the target board KR260 was approximately 55,000 reservoir neurons, as shown in Table I.

B. Power Consumption

We estimated the power consumption for each design. The target board and development environment were identical to those used to evaluate the circuit size. We evaluated the power consumption using Vivado's reported power function based on the post-implementation design. Table III shows the estimated results for $N_r = 4096$. Furthermore, the total power consumption values for $N_r = 128, 256, 512, 1024, 2048$ were 2.848, 2.848, 2.850, 2.852, and 2.856 W, respectively. No significant difference was observed compared with the configuration with $N_r = 4096$. This indicates that increasing the N_r has a limited effect, because the power consumption of the accelerator is a small proportion of the total.

C. Accuracy and Precision

We evaluated the accuracy of the hand-waving action recognition using LUTNet-RC implemented on an FPGA. For the experiment, we used the time-series data of hand-waving actions collected from four subjects for a total of 720 data files. Each file comprised 45 frames. We randomly divided the data into training and test data for each participant, allocating 80% and 20% of the training and test data. This dataset included variations in shooting distance (3, 5, or 7 meter), hand position (left or right), and movement size and speed. It also included data on multiple non-hand-waving movements, such as raising one's hands, sitting down, and walking. These data were used to evaluate recognition accuracy under various conditions. Each data file included input data for the left and right hands, as well as the correct labels. The correct labels were assigned binary classification labels of 1 for hand-waving actions and 0 for non-hand-waving actions. In addition to the LUTNet-RC model, we conducted experiments using the ESN model, which is a general RC model used in previous studies [19]. The same dataset was used for both models to compare their accuracies. The hyperparameters in the RC model were optimized using Optuna [24] over 100 trials. In this study, we evaluated the accuracy and precision of various N_r values for an RC model optimized by Optuna.

Table II shows the experimental results. The results showed that LUTNet-RC achieved higher accuracy and precision than ESN for all N_r.

V. Discussion

A. Accuracy and Precision

In this study, we evaluated the recognition performance of hand-waving actions using accuracy and precision as evaluation metrics. This is because of false positives, in which the wake-up system of a service robot misrecognized *not waving* as *waving*, which directly resulted in decreased usability, such as the wastage of computing resources, due to unnecessary interactions and user discomfort.

Table II shows that LUTNet-RC achieved a better performance than the conventional ESNs in terms of both accuracy and precision. The LUTNet-RC structure quantizes the mathematical ESN model. In general, quantization decreases recognition accuracy. However, in this task, quantization improved performance. There were two main reasons for this improvement. First, the limitation of model expressiveness by quantization may have functioned as regularization, suppressing overfitting to the training data, and improving generalization performance. Second, quantization decreases the sensitivity to input fluctuations and increases robustness against noise and environmental changes in images. Based on these results, the FPGA-based offloading system proposed in this study is considered an effective system with better performance than conventional methods for tasks that require two-class classification and robustness, such as wake-up.

TABLE III

Power Consumption of the KR260 Implementing LUTNet-RC

N_r	Total [W]	Static [W]	Dynamic [W]					
			PS [W]	Clocks [W]	Signals [W]	Logic [W]	BRAM [W]	MMCM [W]
4096	**2.869**	0.301	2.419	0.030	0.008	0.008	0.003	0.097

B. Benefits through FPGA Accelerator

In this study, we proposed a system that offloads recognition processing to an FPGA to reduce the computational load during service robot operation. In laptop-based robot systems that are used in competitions and are highly regarded, even though they only perform basic operations, such as object recognition and self-localization, the resource usage rate of the installed laptop is high: CPU memory occupies 25GB out of 32GB, and GPU memory occupies 10GB out of 16GB. This situation indicates the insufficiency of the system's overall processing power in both the CPU and GPU.

For implementing LUTNet-RC with 55,000 reservoir neurons (the largest possible scale for the edge FPGA KR260), in a laptop-based robot system, 10 GB of memory was required. Therefore, offloading processing tasks to FPGAs is expected to enable parallel processing of tasks that cannot be processed by CPUs or GPUs alone. In the wake-up system targeted in this study, an LUTNet-RC with 55,000 neurons is not necessarily required; nonetheless, FPGAs can process more complex and diverse tasks such as gesture recognition, object recognition, and posture control.

As shown in Table III, the total power consumption of the KR260 implementing LUTNet-RC with 4096 reservoir neurons was 2.869 W, of which 2.419 W, the majority of power consumed, was by the PS, which is the CPU on the SoC-type FPGA. The power consumption of the PL unit was less than that of the PS unit, and there was a negligible increase in power consumption caused by the increase in the N_r. The power consumption of 2.869 W is small compared to the rated power consumption (20 W) of the CPU installed on the laptop connected to the robot, and is expected to reduce overall power consumption of the system. The proposed system can solve the problem of short battery runtimes caused by the high power consumption of laptops, which is a serious issue, particularly for service robots.

VI. Conclusion & Future work

In this study, we proposed a method for offloading hand-waving action recognition processing, a task previously performed by a laptop CPU connected to a conventional service robot, to an FPGA accelerator implementing RC and integrating it into the robot system. The study aimed to improve the efficiency and practicality of the entire system while avoiding excessive dependence on laptops. Through experiments, we confirmed multiple practical values of the proposed system. First, in terms of recognition performance, the LUTNet-RC used in the proposed system outperformed the conventional ESN in terms of both accuracy and precision. This can be attributed to the regularization effect and noise resistance caused by quantization and is extremely important for achieving reliable operation in real-world environments. Furthermore, in terms of scalability and power efficiency, large-scale networks can be implemented without straining the resources of the laptop, and the overall power consumption of the system can be significantly reduced.

This study contributes to solving the fundamental issues faced by service robots, such as limited computing resources and short battery runtimes. These results show that the proposed FPGA offload architecture can aid advanced tasks, such as recognizing more complex human behaviors, with high accuracy and low power consumption. In future work, we plan to conduct further verification and experiments to achieve a system capable of parallel processing of multiple recognition tasks on the same FPGA accelerator. In addition, we aim to extend the FPGA offload architecture to other nonverbal communication methods, such as sign language recognition [25], as well as to wake-up systems using event-based vision sensors [26].

References

[1] Ministry of Internal Affairs and Communications (MIC), "Information and Communications in Japan White Paper 2024: Special Feature II - Coexisting with Evolving Digital Technologies," `https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r06/html/nd132200.html`, 2025.08.15(Accessed), 2024.

[2] T. Ono, D. Kanaoka, T. Shiba, S. Tokuno, Y. Yano, A. Mizutani, I. Matsumoto, H. Amano, and H. Tamukoh, "Solution of World Robot Challenge 2020 Partner Robot Challenge (Real Space)," *Advanced Robotics*, vol. 36, nos. 17–18, pp. 870–889, 2022.

[3] Amazon Science, "Amazon Alexa's new wake word research at Interspeech," `https://www.amazon.science/blog/amazon-alexas-new-wake-word-research-at-interspeech`, 2025.08.07(Accessed), 2022.

[4] J. Qi, L. Ma, Z. Cui, and Y. Yu, "Computer vision-based hand gesture recognition for human–robot interaction: a review," *Complex Intelligent Systems*, vol. 10, pp. 1581–1606, 2024.

[5] E. Wang, J. J. Davis, P. Y. K. Cheung, and G. A. Constantinides, "LUTNet: Rethinking inference in FPGA soft logic," in *Proc. IEEE 27th Annu. Int. Symp. Field-Program. Custom Comput. Mach. (FCCM)*, pp. 26-34, 2019.

[6] H. Jaeger, "The "echo state" approach to analysing and training recurrent neural networks-with an erratum note," *German National Research Center for Information Technology GMD Technical Report*, vol. 148, 2001.

[7] C. Gallicchio, "Euler state networks: Non-dissipative reservoir computing," *Neurocomputing*, vol. 579, pp. 127411–127430, 2024.

[8] K. Yoshioka, Y. Tanaka, and H. Tamukoh, "LUTNet-RC: Look-Up Tables Networks for Reservoir Computing on an FPGA," in *Proc. Int. Conf. Field Program. Technol. (ICFPT)*, pp. 170-178, 2023.

[9] S. Mittal, "A survey on FPGA based accelerators for convolutional neural networks," *Neural Computing and Applications*, vol. 32, no. 4, pp. 1109–1139, 2020.

[10] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Generation Computer Systems*, vol. 97, pp. 219-235, 2019.

[11] K. Yoshioka, P. Allred, T. Barton, B. D. Sahoo, Y.-C. Kuan, S.-H. W. Chiang, and H. Tamukoh, "Reservoir Computing with VCO-Based Spiking Neurons for Regression and Classification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 1–5, 2025.

[12] M. Yamaguchi, Y. Katori, D. Kamimura, H. Tamukoh, and T. Morie, “A Chaotic Boltzmann Machine Working as a Reservoir and Its Analog VLSI Implementation,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, pp. 1–7, 2019.

[13] D. Pramanta and H. Tamukoh, “FPGA Implementation of Pulse-Coupled Phase Oscillators Working as a Reservoir at the Edge of Chaos,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 1–5, 2021.

[14] S. Matsumoto, Y. Ichikawa, N. Kajihara, H. Tamukoh, and T. Morie, “FPGA Implementation for Large-Scale Reservoir Computing Based on Chaotic Boltzmann Machine,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 1–5, 2024.

[15] S. Hirayae, K. Yoshioka, A. Yokota, I. Kawashima, Y. Tanaka, Y. Katori, O. Nomura, T. Morie, and H. Tamukoh, “Enhancing Memory Capacity of Reservoir Computing with Delayed Input and Efficient Hardware Implementation with Shift Registers,” in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, pp. 1–5, 2024.

[16] A. Mizutani, K. Isomoto, K. Yamao, R. Kobayashi, S. Fumoto, K. Arimura, N. Yamaguchi, T. Shiba, K. Kimizuka, Y. Ohno, R. Terashima, H. Yamaguchi, T. Fujino, R. Maruno, W. Yoshimura, K. Mine, T. P. T. Nhan, Y. Yano, Y. Tanaka, T. Nishida, T. Morie, and H. Tamukoh, “Hibikino-Musashi@Home RoboCup@Home DSPL Champion 2024,” in *Proc. RoboCup 2024: Robot World Cup XXVII*, Cham, Springer Nature Switzerland, pp. 461–472, 2025.

[17] K. Isomoto, A. Mizutani, F. Matsuzaki, H. Sato, I. Matsumoto, K. Yamao, T. Kawabata, T. Shiba, Y. Yano, A. Yokota, D. Kanaoka, H. Yamaguchi, K. Murai, M. Kim, L. Shen, M. Suzuka, M. Anraku, N. Yamaguchi, S. Fujimatsu, S. Tokuno, T. Mizo, T. Fujino, Y. Nakadera, Y. Shishido, Y. Nakaoka, Y. Tanaka, T. Morie, and H. Tamukoh, “Hibikino-Musashi@Home 2024 Team Description Paper,” arXiv preprint arXiv:2410.06192, 2024.

[18] M. Quigley, K. Conley, B.P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler and A.Y. Ng, “ROS: an open-source Robot Operating System,” ICRA Workshop on Open Source Software, 2009.

[19] H. Yamaguchi, A. Mizutani, A. R. Syulistyo, Y. Tanaka, and H. Tamukoh, “A Low Computational Cost Hand Waving Action Recognition System with Echo State Network for Home Service Robots,” in *Proc. Int. Conf. Artif. Life Robot. (ICAROB)*, pp. 421-424, 2024.

[20] Y. Ishida, T. Morie, and H. Tamukoh, “A hardware intelligent processing accelerator for domestic service robots,” *Advanced Robotics*, vol. 34, no. 14, pp. 947–957, 2020.

[21] A. Xilinx, “PYNQ,” `https://github.com/Xilinx/PYNQ`, 2025.08.07(Accessed), 2018.

[22] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. A. Sheikh, “Open Pose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, pp. 172-186, 2021.

[23] K. Yoshioka and H. Tamukoh, “A 340-μW TinyML Using LUT-Based Reservoir Computing on Low-Cost FPGAs,” *IEEE Embedded Systems Letters*, vol. 17, no. 5, pp. 357–360, 2025.

[24] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, “Optuna: A Next-generation Hyperparameter Optimization Framework,” in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 2623-2631, 2019.

[25] A. Syulistyo, Y. Tanaka, D. Pramanta, N. Fuengfusin, and H. Tamukoh, “Low-cost computation for isolated sign language video recognition with multiple reservoir computing,” *PLOS ONE*, vol. 20, no. 7, e0322717, 2025.

[26] R. Toyoda, K. Yoshioka, and H. Tamukoh, “Dynamic Threshold Spatial-Temporal Filter on FPGAs for Event-Based Vision Sensors,” in *Proc. IEEE/SICE Int. Symp. Syst. Integr. (SII)*, pp. 854–859, 2025.