

NAVIGATION

Brain-inspired multimodal hybrid neural network for robot place recognition

Fangwen Yu^{1†}, Yujie Wu^{1,2†}, Songchen Ma^{1†}, Mingkun Xu¹, Hongyi Li¹, Huanyu Qu¹,
Chenhang Song¹, Taoyi Wang¹, Rong Zhao^{1,3*}, Luping Shi^{1,3,4*}

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works

Place recognition is an essential spatial intelligence capability for robots to understand and navigate the world. However, recognizing places in natural environments remains a challenging task for robots because of resource limitations and changing environments. In contrast, humans and animals can robustly and efficiently recognize hundreds of thousands of places in different conditions. Here, we report a brain-inspired general place recognition system, dubbed NeuroGPR, that enables robots to recognize places by mimicking the neural mechanism of multimodal sensing, encoding, and computing through a continuum of space and time. Our system consists of a multimodal hybrid neural network (MHNN) that encodes and integrates multimodal cues from both conventional and neuromorphic sensors. Specifically, to encode different sensory cues, we built various neural networks of spatial view cells, place cells, head direction cells, and time cells. To integrate these cues, we designed a multiscale liquid state machine that can process and fuse multimodal information effectively and asynchronously using diverse neuronal dynamics and bioinspired inhibitory circuits. We deployed the MHNN on Tianjic, a hybrid neuromorphic chip, and integrated it into a quadruped robot. Our results show that NeuroGPR achieves better performance compared with conventional and existing biologically inspired approaches, exhibiting robustness to diverse environmental uncertainty, including perceptual aliasing, motion blur, light, or weather changes. Running NeuroGPR as an overall multi-neural network workload on Tianjic showcases its advantages with 10.5 times lower latency and 43.6% lower power consumption than the commonly used mobile robot processor Jetson Xavier NX.

INTRODUCTION

Place recognition has gained increasing attention in the robotics community to help robots understand the spatial characteristics of the world. It refers to the problem of deciding whether a place has been visited before, and, if it has been visited before, which place it was (1). Many recent studies have achieved great progress in improving the performance of place recognition with a single image or jointly with image, geometric, and semantic information (1–5). However, place recognition in natural environments remains a huge challenge because of rapid environmental changes and stringent requirements for power, computing, and latency caused by the limited resources of robots. Three crucial challenges still retard the development of place recognition, including how to sense places reliably, represent places robustly, and match places efficiently.

To address these challenges, roboticists have turned to nature for inspiration. At present, there are two main approaches: designing biologically inspired visual place recognition (VPR) methods (6–12) and using neuromorphic sensors, such as event cameras (13–15), to improve the precision and robustness of place recognition (16–22). Despite the advancements brought by various biologically inspired approaches, most of them still face three main problems. First, they often fail to address changing physical locations or

viewpoints when only a single sensory cue is used. Second, the place representation captured by a single conventional frame-based camera or neuromorphic sensor leads to the loss of a large amount of information in varying environments. Third, the computational cost scales linearly with the dataset size. In large-scale wild environments, the high computational load causes delays in place matching, degrading the real-time performance of robots.

In contrast, humans and animals demonstrate remarkable place recognition capabilities, robustly identifying places in large three-dimensional environments (23–26). They can reliably sense, robustly represent, and efficiently recognize places using an internal conjunctive representation of spatial view cells (27, 28), auditory view cells (3, 29–31), olfactory view cells (32–34), place cells (35, 36), head direction cells (37, 38), grid cells (39, 40), and time cells (41–43). As illustrated in Fig. 1A, when people recognize a place, different sensory cues could activate multiple types of sensory cells (28, 44, 45). These cells can provide a spatiotemporal representation of a place that extends along a continuum, with each moment containing information about the past, present, and future (26), facilitating the precision and robustness of place recognition. However, such hybrid spatiotemporal continuity of perception and encoding is still lacking in current methods of place recognition. Combining multimodal sensing and the property of hybrid spatiotemporal computing may offer perspectives for the development of place recognition, but this has yet to be explored.

Neuromorphic computing provides an emerging hardware platform for the spatiotemporal computing paradigm using neural principles of integrating information processing and memory in the brain, demonstrating advantages of low power consumption, multiple network parallelism, and low delay inferences (46–58). One of the recent advances in this field is the Tianjic chip, which

¹Center for Brain-Inspired Computing Research (CBICR), Optical Memory National Engineering Research Center, and Department of Precision Instrument, Tsinghua University, Beijing 100084, China. ²Institute of Theoretical Computer Science, Graz University of Technology, Graz, Austria. ³IDG/McGovern Institute for Brain Research, Tsinghua University, Beijing 100084, China. ⁴THU-CET HIK Joint Research Center for Brain-Inspired Computing, Tsinghua University, Beijing 100084, China. *Corresponding author. Email: lpshi@tsinghua.edu.cn (L.S.); r_zhao@tsinghua.edu.cn (R.Z.)

†These authors contributed equally to this work.

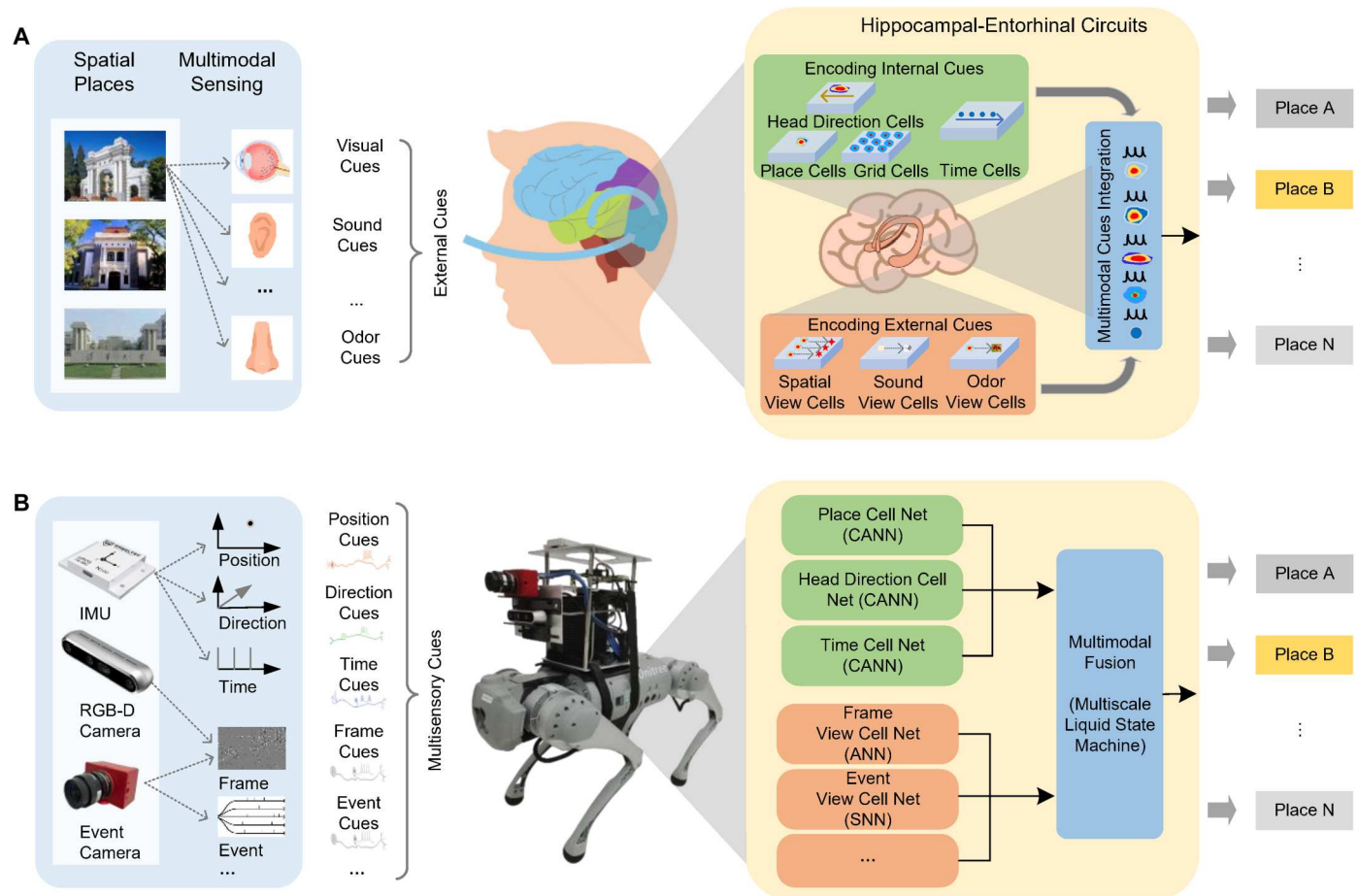


Fig. 1. The mechanism of place recognition in humans and robots. (A) Humans can recognize a place according to multimodal sensory cues with spatiotemporal continuity. The place is encoded by external sensory cells and internal conjunctive spatiotemporal cells. The multimodal sensory cues could activate these cells together. The conjunctive firing pattern can encode and recall a specific place uniquely. (B) The robot can recognize a place with a brain-inspired place recognition system. Similarly, the robot can obtain sensory cues of the environment using multimodal sensors. An MHNN model deployed on a neuromorphic computing chip was used to encode these sensory cues, which can recognize places with a multiple–spatiotemporal scale fusion approach robustly.

provides a unified platform that supports versatile neural network models with high efficiency. This platform can execute artificial neural networks (ANNs) and spiking neural networks (SNNs) independently or in a hybrid manner (46–48, 51), providing a strategic opportunity to overcome the bottlenecks of conventional VPR.

Here, we report a neuromorphic general place recognition system, dubbed NeuroGPR, that emulates the neural mechanisms of multimodal sensing, coding, and computing, enabling robots to recognize places in natural environments robustly and efficiently. We addressed several key challenges in the design of NeuroGPR.

First, to enhance the reliability of a single sensor, we integrated multiple conventional sensors and neuromorphic sensors in our systems for multimodal sensing. This enables the robot to effectively capture and process multiscale spatiotemporal information by emulating multimodal sensing mechanisms of the brain.

Second, to encode and fuse multiscale, multidimensional information from different sensors, we designed a multimodal hybrid neural network (MHNN) model and a multiscale liquid state machine (MLSM) model for asynchronous fusion of multimodal information. The MLSM is built on heterogeneous spiking neurons

for information processing, which emulates the diverse neural information processing mechanisms of the brain and projects the multimodal information into a high-dimensional sparse representation space using local inhibitory circuits. By optimizing the network parameters through gradient descent techniques, MLSM can effectively fuse the multimodal information and enable the robot to work robustly even when parts of the modal information are missing or changed.

Third, we used the hybrid neuromorphic chip Tianjic to efficiently execute multiple cross-paradigm networks in parallel. The Tianjic chip allows the robot to concurrently realize multiple neural networks with various scales, precisions, and structures using independently configurable function cores. Its flexible data movement mechanism and adjustable timing schedules enabled efficient data interactions and signal conversions between different networks. We deployed the MHNN on Tianjic to perform place recognition.

Last, we developed a quadruped robot system equipped with multimodal sensors and hybrid neuromorphic chips to evaluate the NeuroGPR system (Fig. 1B). We tested NeuroGPR in natural

environments and demonstrated that it outperformed the state-of-the-art conventional techniques in both indoor and outdoor environments, exhibiting robustness to environmental uncertainties.

RESULTS

Multimodal hybrid neural network

The MHNN is composed of convolutional neural networks (CNNs), SNNs, and continuous attractor neural networks (CANNs) for representing multisensory cues and an MLSM for fusing multimodal information. The overall structure of the MHNN is shown in Fig. 2A. Specifically, we adopted versatile neural encoding strategies, such as rate-based coding and Gaussian population coding, to encode heterogeneous sensory cues. Different network modules processed distinct features and fed multimodal information into the MLSM. All modules were ultimately integrated within a unified optimization framework, which was optimized using the backpropagation-through-time (BPTT) algorithm.

Considering the spatiotemporal multimodality of the external input data from different cells, we used the representative CNN, SNN, and CANN for modeling the spatial view cells, place cells, head direction cells, and time cells, thus encoding different types of inputs. Specifically, because nonspiked CNNs are generally adept at processing information with high spatial complexity, we adopted it to model spatial view cells for processing frame-based visual cues. Spiking networks mimic biological spike-based information processing mechanisms that can naturally extract temporal correlations and are conducive to processing event-driven spatiotemporal data streams. We adopted an SNN model to model the function of spatial view cells for event-based visual cues. In addition, place cells and head direction cells have been widely studied using neuroscience-inspired CANNs, providing generic and simple models for low-dimensional spatiotemporal continuum features. Because the robot needed to consider its local relative position, direction, and time information when it moved through

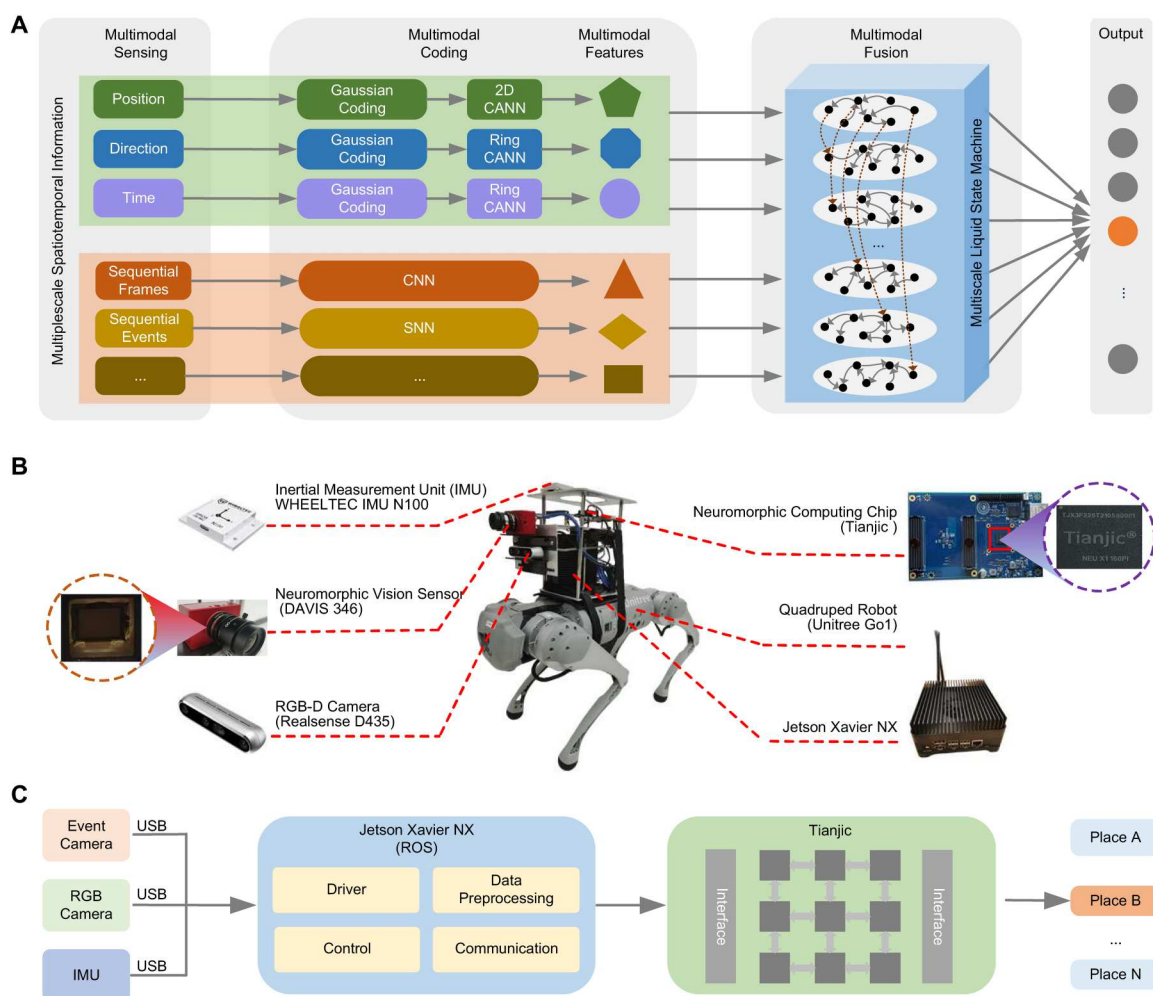


Fig. 2. The NeuroGPR system architecture. (A) The MHNN model is composed of a CNN, an SNN, and three CANNs. The multimodal features are extracted from sensory cue representations, which are combined by a multiple-spatiotemporal scale fusion approach using the MLSM. (B) Key components of the NeuroGPR system. The system comprises multimodal sensors, the neuromorphic computing chip Tianjic, and a quadraped robot. (C) The pipeline of the NeuroGPR system. The host computing platform is dedicated to sensor data acquisition, data preprocessing, robot control, and data communication. The MHNN is mapped and deployed on Tianjic, which recognizes places on the basis of the inputs of query data.

space, we introduced three CANNs to model the functions of place cells, head direction cells, and time cells, respectively.

To integrate these encoded data by different modules, we built the MLSM model with a liquid pool of differentiable spiking neurons with the membrane dynamics, adaptive threshold, and adaptive currents. As shown in Fig. 2A, various sensory information was projected into different regions of the liquid machine. The local inhibitory circuits and lateral inhibitions controlled the information representation and interaction for each modality of information. We used the BPTT algorithm to optimize the synaptic weights and hyperparameters of neuron dynamics. Last, we trained a classifier to decode spike signals from the pool and produce the recognition results.

System architecture

We built the NeuroGPR system on the basis of a quadruped robot to perform place recognition tasks in real time. As shown in Fig. 2B, we deployed a neuromorphic dynamic vision sensor (DVS); an Intel RealSense RGB-D camera; a WHEELTEC IMU N100; a host computing platform, Nvidia Jetson Xavier NX (known as Xavier NX); and a Tianjic computing platform on the robot. Ubuntu 18.04 and Robot Operating System Melodic (59), were installed on Xavier NX, which can preprocess sensor data and control the robot. A toolchain of mapping, compiling, and execution of Tianjic was used to process the NeuroGPR model. The pipeline of the NeuroGPR system is shown in Fig. 2C. The DVS, RGB-D camera, and IMU were connected to Xavier NX. Tianjic obtained the sensor data from Xavier NX. The parallel network data flow of the MHNN was deployed on the Tianjic chip for recognizing places according to the inputs.

We leveraged Tianjic to support the MHNN for place recognition. Tianjic can support the parallel execution of multiple cross-paradigm hybrid neural networks and has a unified, configurable, and scalable architecture with reconfigurable units supporting rich coding schemes and hybrid models. Each functional core can be configured independently, enabling the chip to concurrently implement multiple neural networks with various neural types, scales, precisions, and structures. Moreover, it provides flexible data movement mechanisms and adjustable timing schedules, enabling efficient data interactions and signal conversions between different networks.

In the following sections, we evaluated NeuroGPR's real-time performance from computing efficiency, power consumption, and recognition accuracy. We also compared NeuroGPR with the state-of-the-art biologically inspired VPR approaches. Last, we investigated the effects of different network modules and the robustness of environmental changes.

Performance evaluation

We validated the performance of the NeuroGPR system on a robot with different scenarios and conditions. As illustrated in Fig. 3, the test scenarios were set up in a room, a long corridor, and a wild forest. In different locations or floors, the appearances of the environment were very similar from the same or opposite viewpoints, and the light changed a lot in different areas. When the robot moved into the dark area, the image captured from the conventional camera became fully dark, indicating that useful information was lost, as shown in Fig. 3D. Furthermore, to add uncertainty to the environment, we designed the scenario so that some people

suddenly entered or left the corridor without notice. These changing factors can help us evaluate the robustness of the NeuroGPR system in natural environments. In addition, we tested the system in a forest environment, as shown in Fig. 3F. The robot vibrated when moving on the cobble road, which led to motion blur of RGB images and noise in event camera data. The light changed at different times, leading to large changes in the RGB image that were captured in the same place.

Computing performance

We conducted a computing performance of the MHNN model on the Tianjic chip and compared the performance with those of other computing platforms, including Xavier NX and Nvidia Jetson AGX Orin (known as AGX Orin), which are commonly used for mobile robots and edge intelligent applications. Both of them have integrated central processing units (CPUs), graphics processing units (GPUs), and peripheral interfaces. The AGX Orin with higher computing power was released later than the Xavier NX. Because of the limited resources on the hardware systems, we simplified the MHNN for the deployment and ensured consistent network models across all comparison platforms.

Computing latency is the key performance indicator for place recognition in robots. However, in the scenario of mobile robotics, standardized metrics for evaluating the real-time processing capability of multi-neural network (multi-NN) models like the MHNN are now lacking. To assess multi-NN performance reasonably, we considered the computing latency performances of these platforms in two aspects: single-NN workloads and multi-NN workloads. To demonstrate different features of the MHNN model, the single-NN workloads evaluated the average timing duration for processing inputs (after preprocessing) from any sensor and updating the MLSM until a valid output was obtained. In real-world and practical applications, robots usually receive input data from various sensors simultaneously, necessitating concurrent execution of different neural networks by the processor. Thus, the multi-NN workloads evaluated the task-level parallelism performance of computing hardware.

We executed independent operations and performance tests of three groups of single-NN workloads on multiple hardware platforms, namely, CNN + MLSM, SNN + MLSM, and CANN + MLSM (corresponding to the three different sensor inputs exactly). These tests enable simple estimation of the computing performance of a hardware platform when operating different types of networks. As shown in Fig. 4A, Tianjic demonstrated competitive real-time performance in three different single-NN workloads. When processing the CNN + MLSM on Tianjic, the average computing latency was decreased by 93.4 and 87.8% compared with Xavier NX and AGX Orin, respectively. For the SNN + MLSM, the average computing latency of Tianjic was decreased by 87.3 and 77.0% compared with another two processors, respectively. We carried out parallel optimization for GPUs according to their architectural features to achieve a fair and objective evaluation: continuously accumulate nine frames of DVS input and then send them to SNN for processing. In this case, the GPU only needs to run the MLSM once every nine frames. However, because Tianjic embodied a data-driven execution pattern, the computing latency was calculated by adding the nine frames' processing time of both the SNN and the MLSM. When processing the CANN on GPUs, the computing latency of Xavier NX was about 692.4 times the number of Tianjic, and AGX Orin was about 304.1 times. Through analysis,

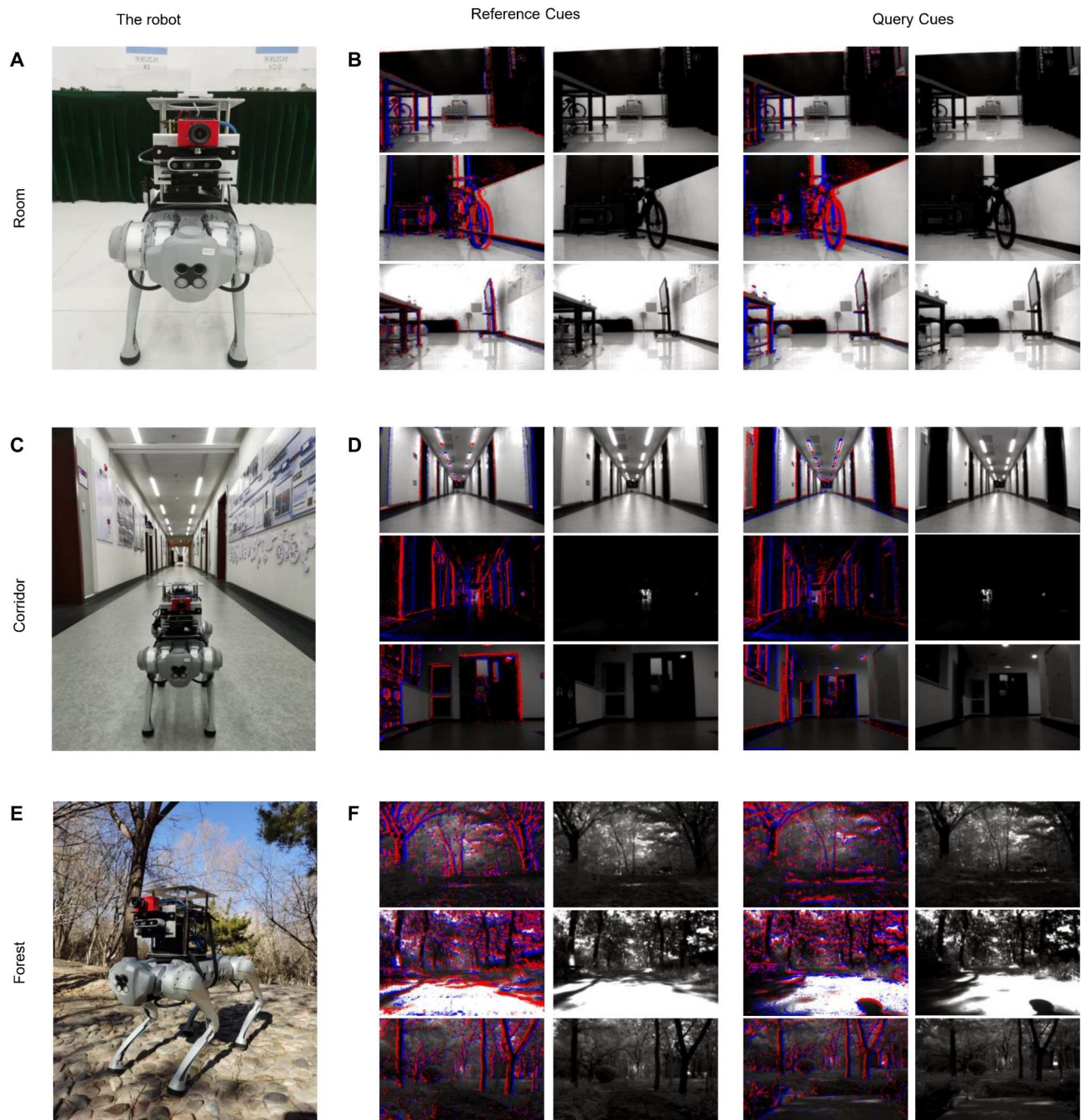


Fig. 3. Snapshots of the experimental environments. Snapshots of the robot experiment in the room (A), corridor (C), and forest (E) environments. Snapshots of the reference cues for training and the query cues for testing in the room (B), corridor (D), and forest (F) environments.

the computation for CANN involved solving ordinary differential equations (ODEs), but running CANN on GPU requires more intensive computation resources. Alternatively, Tianjic discretized and iterated the ODE solving process with time steps and adopted a lookup table to realize nonlinear function solving, leading to the improvement in performance. For objectivity and fairness, we also deployed the CANN on the CPUs of the two processors, and performance improved. In this case, the computing latency of Xavier NX was about 34.2 times the number of Tianjic,

and AGX Orin was about 16.2 times. Thus, Tianjic demonstrated competitive real-time performance in single-NN workloads.

We further evaluated the performance in the processing of multi-NN workloads. GPUs typically adopt the single-instruction multiple-thread architecture and rely on the multiprocessing of the operating system to implement the multi-NN scheduling and execution, leading to an extra delay in the task switch. Conversely, Tianjic's many-core multiple-instruction multiple-data architecture enabled spatial multitasking and pipelining, allowing each neural

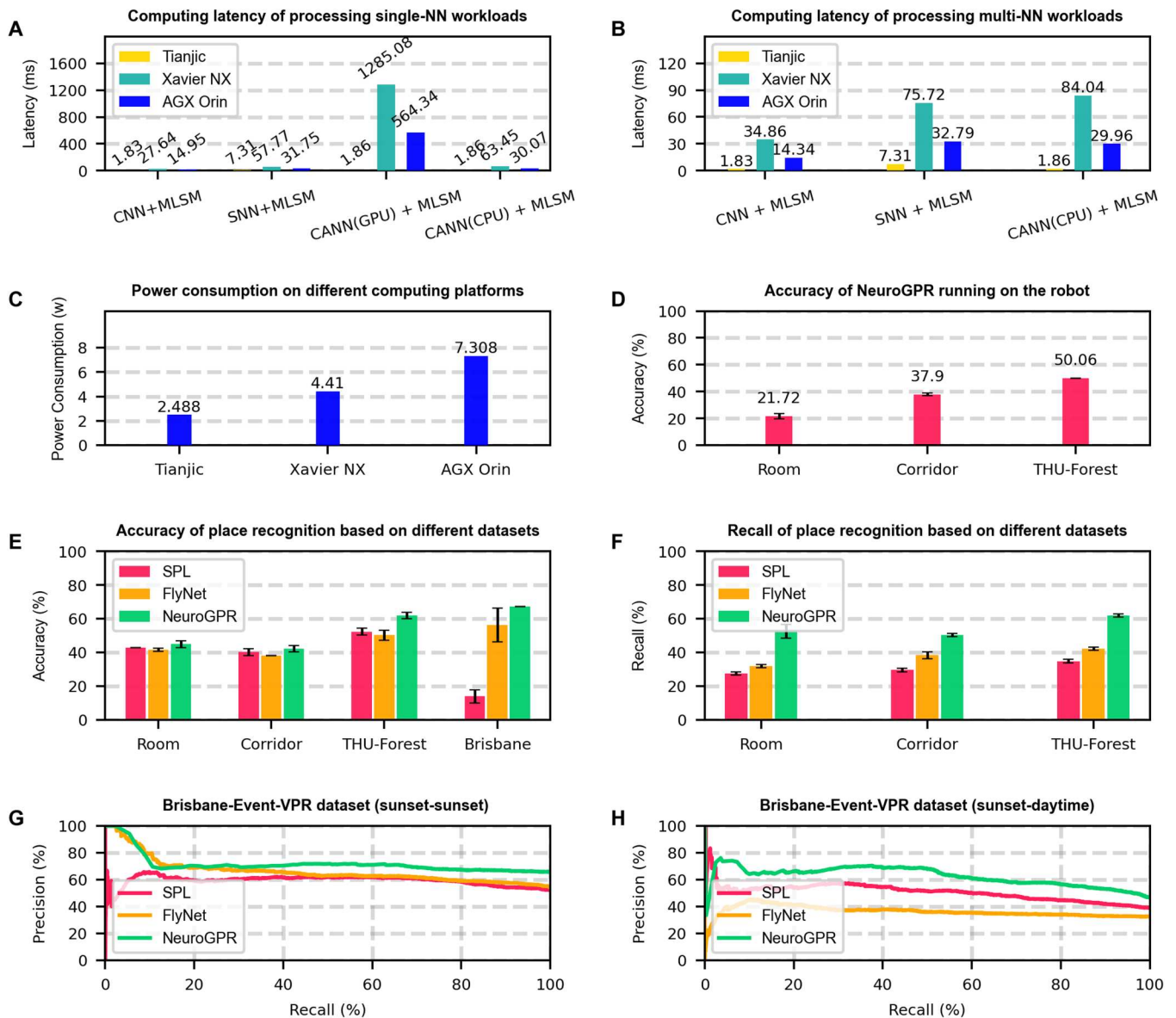


Fig. 4. The experimental results of computing efficiency and accuracy. Computing latency of processing single-NN workloads (A) and multi-NN workloads (B). (C) The power consumption of processing MHNN under the multi-NN workloads test. (D) Accuracy of NeuroGPR running on the robot. The accuracy (E), the recall (F), and the precision-recall (G and H) curves of place recognition based on different datasets. The mean errors and SDs are based on the results obtained in five experiments.

network to be deployed in parallel on separate core clusters and alleviating the resource preemption and switching delay. Consequently, when the computing and storage resources were sufficient, the computation delay of the multi-NN workloads matched that of the single-NN workloads. The aforementioned three single-NN workloads were submitted to Xavier NX and AGX Orin concurrently via multiprocessing, and the average latency of each group of workloads was tested. As shown in Fig. 4B, the computing delay of each workload was increased compared with the independent execution in most cases. Xavier NX could only achieve concurrent execution of two processes at a time because of bandwidth limitation. Thus, we conducted four sets of tests for Xavier NX to examine a wider range of concurrent executions. The computing latency of Xavier NX was the average value of four sets of

tests (see fig. S10 for details). For execution of multi-NN workloads, when processing on Tianjic, the average computing latencies for three groups of workloads were still able to be decreased by 90.3 to 97.8% and 77.7 to 93.8% compared with when they ran on Xavier NX and AGX Orin, respectively. At this point, we got comparative data by deploying CANN on their CPUs. If CANN is deployed on the GPUs of Xavier NX and Orin, then Tianjic can achieve hundreds of times (about 303.5 times) the performance improvement data. Under multi-NN workloads evaluation, the computing latency for the overall MHNN could be equivalent to the largest latency value in all neural networks processing. Thus, the computing latency for the overall MHNN model processing on Tianjic, Xavier NX, and AGX Orin was 7.31, 84.04, and 32.79 ms, respectively. In other words, compared with Xavier NX and AGX

Downloaded from https://www.science.org at The Hong Kong University of Science and Technology (Guangzhou) on May 25, 2026

Orin, Tianjic achieved 10.50 and 3.49 times improvement in real-time processing performance for the overall MHNN model, respectively.

Power consumption is another key measurement for the long-term operation of robots, particularly in the wild environment. We further evaluated the average power consumption of the MHNN model under the test of multi-NN workloads on different processors. The MHNN running on Tianjic consumed 43.6% less power than on Xavier NX and 66.0% less than on AGX Orin (GPU specific), as shown in Fig. 4C. Because Xavier NX had difficulty with submitting three neural network processes for concurrent running, to obtain the power performance of multi-NN workloads, we had to integrate three neural networks into one process for testing. Because of insufficient utilization of hardware resources, this method led to the decline of power consumption on this processor. Compared with conventional processors, Tianjic introduces a multiple primitive execution mechanism for the many-core neuromorphic architecture, enabling each functional core to execute with time-division multiplexing and reuse. By integrating the asynchronous execution mechanism of the spatial grouping of functional cores, Tianjic can support spatiotemporal hybrid mapping efficiently. Specifically, layers in a computing

task can be divided into different layer groups. Different layer groups can be deployed to different temporal core clusters in space for asynchronous execution to leverage the advantage of parallel pipelined execution of the many-core neuromorphic chips. The layer in each layer group is also completed in parallel through multiple functional cores with the maximized utilization of computing and storage resources in the local functional core through temporal mapping. Therefore, the hardware utilization and throughput can be enhanced (48, 51). In the mobile robot scenarios, the frame rate of the sensor directly affects the throughput of the system, so we would not prioritize the evaluation of the computing throughput.

Accuracy evaluation of place recognition

We next demonstrated the practical performance of NeuroGPR in different environments using the accuracy metric, which evaluates how many places the robot can recognize correctly in all queries. Figure 4D shows the accuracies of NeuroGPR in the room, corridor, and forest environments, indicating high recognition accuracies and stable performance in different practical environments.

To further evaluate NeuroGPR, we compared its performance with those of two state-of-the-art biologically inspired place recognition methods, FlyNet and SPL, in different environments. We

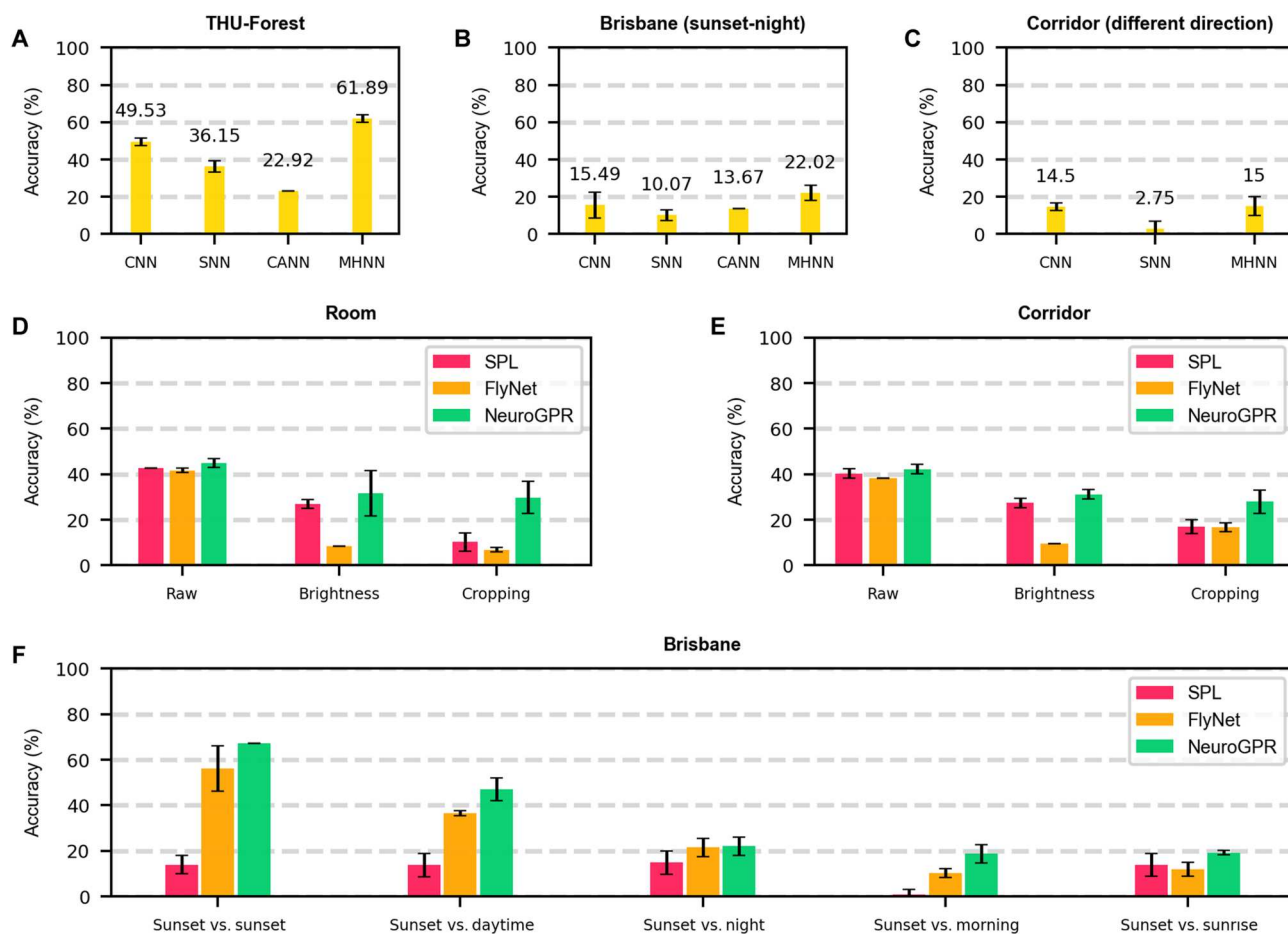


Fig. 5. Experimental results of robustness to environmental changes. Influence of different modules on the recognition accuracies of the THU-forest dataset (A), the Brisbane-Event-VPR dataset (B), and the corridor dataset (C). Evaluation of the recognition accuracies of different models on the room dataset (D), the corridor dataset (E), and the Brisbane-Event-VPR dataset (F). The mean errors and SDs are based on the results obtained in five experiments.

collected the datasets in large-scale indoor and outdoor environments. The properties of these datasets are listed in table S2, and the relevant snapshots and distance matrices are shown in figs. S1 to S8.

In Fig. 4E, we compared the place recognition accuracy of different methods in various environments. In the room, corridor, forest, and road environments, NeuroGPR outperformed SPL and FlyNet. Figure 4F demonstrates that NeuroGPR also had higher recall rates than SPL and FlyNet in three different environments. Figure 4 (G and H) shows the precision-recall curves of the three methods on the Brisbane-Event-VPR dataset, where the sunset-1 data were used to train these models and the sunset-2 data (Fig. 4G) and the daytime data (Fig. 4H) were used to test each model. The precision-recall curves demonstrate that NeuroGPR achieved more robust and precise place recognition than SPL and FlyNet in different weather and light conditions.

Influence of different network modules

We investigated the influence of the single network module on the performance of NeuroGPR in various conditions. We designed three control models by selectively turning off parts of information processing modules, including a single CNN module, an SNN module, and three CANN modules. In these models, the MHNN combines the information from the remaining open modules for place recognition. We evaluated the function of different modules on various datasets, including the Tsinghua University (THU)-forest, Brisbane-Event-VPR, and corridor datasets. The firing patterns of the place cells and head direction cells represented by the CANN modules are provided in fig. S9.

As shown in Fig. 5A, the MHNN outperformed other modules on the THU-forest dataset, achieving higher accuracies than the CNN, SNN, and CANN modules. Similarly, Fig. 5 (B and C) demonstrates that NeuroGPR achieved higher accuracies than single-module models on the sunset-night data in both the Brisbane-Event-VPR dataset and the corridor datasets. The corridor dataset contained data where the robots navigated in the same place but with the opposite direction. The high performance of NeuroGPR shown in Fig. 5C indicates that the position and direction information can help the NeuroGPR system eliminate errors caused by the scene similarity from different locations or different views of the same or different floors. Collectively, our results indicate that NeuroGPR can integrate multisensory cues for comprehensive decision-making, provide high recognition accuracy, and enable more flexible use of different networks to adapt to changing environments.

Robustness to environmental changes

Environmental conditions can change rapidly and unpredictably in natural settings. We next evaluated the performance of NeuroGPR on the three typical environmental changes: light, weather, and appearance. To simulate the natural environments, we added noise to the room and the corridor datasets by reducing the image brightness by 50% and cropping the image size from 346 pixels by 260 pixels to 128 pixels by 128 pixels. In addition, we directly used the Brisbane-Event-VPR dataset at different times and weather conditions to test the system's robustness.

As shown in Fig. 5 (D and E), NeuroGPR exhibited higher performance compared with SPL and FlyNet in changing environments. Even after reducing image brightness or cropping the image size, the performance of NeuroGPR was not affected. The other modules continued to provide stable and robust performance when frame-based signals deteriorated. Figure 5F shows that

NeuroGPR outperformed SPL and FlyNet in different lighting and weather conditions, highlighting its strong robustness. In all conditions, NeuroGPR achieved higher accuracy compared with SPL and FlyNet. These results indicate that NeuroGPR holds promise for robust operation in uncertain and changing environments.

DISCUSSION

We have presented a brain-inspired general place recognition system that performs multimodal sensing, coding, and computing, enabling a mobile robot to recognize places in natural environments. Unlike conventional and biologically inspired place recognition approaches, we leveraged multimodal sensing, an MHNN, and a hybrid neuromorphic chip to process the multimodal and multiple-spatiotemporal scale information for place recognition. We evaluated the NeuroGPR system on the basis of a quadruped robot in real-time and large-scale datasets offline in both indoor and outdoor natural environments. The experimental results showed that NeuroGPR not only was competitive in performance but also demonstrated capabilities with high robustness, low latency, and low power consumption.

To improve the performance of a single sensor, we integrated the conventional and neuromorphic sensors and emulated multiple neural coding strategies. For encoding and fusing multiscale and multidimensional information, we introduced the MHNN and proposed the MLSM model. The MLSM leveraged the sparse neuron activities for representing different sensory information and enabled the model to process multimodal information asynchronously. We demonstrated the effectiveness and robustness of the MLSM in multiple natural environments.

We also leveraged the neuromorphic Tianjic chip to implement the MHNN efficiently. The Tianjic chip offers concurrent realization of multiple neural networks with various scales, precisions, and structures. Our results on the chip demonstrated that the flexible data movement mechanism and adjustable timing schedules enable efficient data interactions and signal conversions between different networks, leading to improved computing efficiency and reduced power consumption compared with conventional hardware systems.

Collectively, the proposed NeuroGPR system achieved competitive performance, high robustness, low latency, and low power consumption compared with conventional and biologically inspired place recognition approaches. We believe that our work offers insights into how neuromorphic techniques can improve the capabilities of robots. Future work will focus on integrating NeuroGPR into the next generation of fully neuromorphic Simultaneous Localization and Mapping (SLAM) and navigation systems for brain-inspired robotics (60), which will allow autonomous robots to operate in natural environments with high robustness, high computing efficiency, and low power consumption.

MATERIALS AND METHODS

Implementation of MHNN

The MHNN comprises five front-end network modules (including CNN, SNN, and three CANN modules) for processing distinct multimodal sensory input, the MLSM for integrating the processed multimodal information, and a linear decoder for outputting

results. We elaborate on each module of the MHNN model and provide the training details of the overall model. The detailed parameters are listed in table S1.

CNN module

We used pretrained network models for the CNN module. To meet the hardware requirement, the pretrained ResNet50 was simplified and deployed on Tianjic chips. To compare with other algorithms, the pretrained MobileNet-V2 was adopted for demonstrating the algorithmic performance on three datasets. For each frame-based sensory information, the results of the CNN module were incorporated into the network, and the activation in the last hidden layer was taken as the output for the MLSM. In our experiment, the CNN module was fixed without further fine-tuning.

SNN module

An eight-layer convolutional SNN was established with the structure (Input-AP2-64C3P3-128C2P3-256C2P3-256C2P3-AP2-FC512-FC512-FC200) to process event-based information, where AP2 represents the average pooling with a kernel size of 2, C3P3 represents the convolutional layer with a convolutional kernel size of 3 and a padding size of 3, and FC represents the fully connected layer. The leaky integrate-and-fire (LIF) model was adopted to process event-driven spike signals with the following dynamics:

$$\begin{cases} \tau_{u1} \frac{du}{dt} = -(u - u_r) + W^1 I, \\ s(t) = H(u(t) - v_{th}) \end{cases}$$

where τ_{u1} represents the membrane potential time constant, u_r represents the resting potential, $H(x)$ represents the Heaviside function, W^1 represents the input weights, and v_{th} denotes the neuronal firing threshold. Equation 1 describes that the membrane potential receives the weighted sum of input currents I with time. When the membrane potential exceeds the given threshold v_{th} , the LIF neuron fires a spike and resets the membrane potential to u_r ; otherwise, the membrane potential accumulates to the next time step. In our experiment, the event-based data were directly input into the module, and the module output was sent to the MLSM for further processing. We used the surrogate function techniques (61, 62) to approximate the derivative of the spiking function $H(x)$.

CANN module

We established a position-based two-dimensional CANN with 64 neurons by 64 neurons to process the position information. We took the Gaussian encoding method and constructed one neuron population with 64 neurons by 64 neurons, where each neuron has its independent receptive field. In this manner, the position information was encoded by the neuron population activities. We followed the work (63) to build the CANN modules with continuous attractor dynamics. Typically, the encoded spike signals were continuously input into the network. The continuous attractor dynamics was evolved following the equation below with a 0.05-ms simulation time step:

$$\begin{cases} \tau_{u2} \frac{du}{dt} = -u(x, t) + \alpha \int J(x, x') r(x', x) dx' + I^{ext}(x, t) \\ J(x, x') = J_0 e^{-\frac{(x-x')^2}{2}} \\ r(x, t) = \frac{u(x, t)^2}{1 + \alpha \int u(x, t)^2 dt} \end{cases}$$

where τ_{u2} denotes the synaptic time constant, α denotes the connection density, and $J(x, x')$ is a function of $(x - x')$ that represents the connection strength from neuron x to x' and is normalized by the

connection constant J_0 . The term $I^{ext}(x, t)$ denotes the afferent spike trains by Gaussian encoding. Last, the normalized neuron activities $r(x, t)$ at the last time step are read as the output.

We established a time-based one-dimensional CANN with 128 neurons to process the time information and used the Gaussian population coding to encode the time information. When the robot traversed the road, the relative motion time from a specific place, for example, the starting place, was taken as the input of the time cell model. We also established a direction-based one-dimensional CANN with 128 neurons to process the head direction information and used the Gaussian population coding to encode the head direction information. When the robot suddenly stopped, the relative motion time was kept constant. The motion state was detected on the basis of the odometry or IMU information. We modeled the CANN models of time cells and head direction cells in a similar manner as described in Eq. 2.

Multiscale liquid state machine

The MLSM constructs the liquid pool using the differentiable spiking neuron models. This neuron model is based on a specific type of generalized LIF model known as the GLIF3 model (64), which can mimic hundreds of different cell types with good reproducibility for diverse biological spiking timing features. The MLSM introduces the learnable hyperparameters in the dynamics of the membrane potentials u , input current \bar{I} , and thresholds \bar{v}_{th} , allowing different neurons to process temporal information with varying time scales.

To explore the optimal hyperparameters of the GLIF3 model for place recognition, we adapted the expression of GLIF3 and applied the BPTT for optimizing the hyperparameters of neuronal dynamics and synaptic weights. To make the GLIF3 differentiable, we reformulated the original neuron models using the eligibility trace (65), which models the dynamics regarding the membrane potential, synaptic current \bar{I} , and adaptive threshold \bar{v}_{th} in an explicit iterative expression. The dynamics can be formalized as

$$\begin{cases} \tau_{u3} \frac{du}{dt} = -(u(t) - u_r) + W^2 \bar{I}(t) + Vs(t - \delta t) - s(t - \delta t)\eta \\ \tau_I \frac{d\bar{I}(t)}{dt} = -\bar{I}(t) + \alpha_I x(t) \\ \tau_{th} \frac{d\bar{v}_{th}(t)}{dt} = -(\bar{v}_{th}(t) - \bar{v}_0) + \alpha_{th} x(t) \\ s(t) = LocalWTA(H(u - \bar{v}_{th})) \end{cases}$$

where τ_{u3} denotes the membrane potential time constant, δt denotes the time delay, \bar{v}_0 denotes the resetting values of the threshold, η denotes a resetting constant, V represents sparse lateral connections, W^2 denotes the afferent synaptic weights, and $x(t)$ denotes the output of the front-end modules. τ_I and τ_{th} control the decay rates of each trace, and α_I and α_{th} control the learning rates of different traces. The local winner-take-all (LocalWTA) refers to the local lateral inhibition (66) to ensure that, in a local neuron pool, only a small number of neurons with the largest membrane potentials can fire spikes.

At each simulation time step, different types of sensory information were projected to different nonoverlapping regions of liquid pools. We applied the local WTA mechanism for each region and introduced sparse lateral connections among different regions. Specifically, by applying local WTA, 20% of the neurons with the highest membrane potential are allowed to fire spikes, and other neurons are inhibited by resetting the membrane potential to the resting values. This competition mechanism encourages sparse

information representation for each type of modal information. In addition, random sparse lateral connections with a connection probability of 20% established interregional connections for spike signal exchange, which encouraged information association between different types of information, providing modality completion capabilities for robust place recognition. Last, a linear classifier was used to read out the mean population activity of the liquid pool and produced the final output.

In our experiment, we randomly initialized decay rates and learning rates of the eligibility traces from a uniform distribution $U_{[0,1]}$ to simulate the multiscale dynamics characteristic of biologically heterogeneous cell types. We optimized the parameters of the MLSM by minimizing the cross-entropy loss and adopted an adaptive moment estimation optimizer (67) to accelerate training. Details of the parameter settings are provided in table S1.

Deployment on Tianjic

We developed a hierarchical compiler and simulation toolchain to automatically deploy the MHNN on the Tianjic chip. The Tianjic compiler included three parts: transformer, mapper, and the code generator. The transformer was responsible for the operator conversion, optimizing the computation graph model for specific primitives and hardware precision requirements in Tianjic. It completed the algorithm quantization, converted the algorithm framework into hardware operators, and lastly output the computation graph represented by the Tianjic hardware primitives. Taking the chip's timing mechanism (68) and hardware constraints into account, the mapper generated the position and timing distribution of all task nodes to determine the tasks processed by each functional core in a time phase (the minimum execution time unit). The output of the mapper was a computing graph with mapping information. Through the code generator, we obtained the resource utilization report of the network and the executable file on Tianjic for the chip configuration. Last, the executable file was downloaded into the chip for execution, where the latency and power consumption of network execution were measured accurately.

Performance measurement setup

Multi-NN tasks bring out challenges for meaningful hardware performance evaluation. As a kind of neuromorphic chip, Tianjic has the feature of near-memory computing and massively fine-grained parallelism. Therefore, for fair performance evaluation, we have excluded the latency of data loading in Xavier NX and AGX Orin. Because our evaluation was specifically centered around the parallel processing capabilities of the MHNN, the multimodal neural networks, we chose to primarily compare the performance of Tianjic with those of GPUs in Xavier NX and AGX Orin. Here, all of the results were based on on-chip evaluation, including sample-wise computing latency (single-NN and multi-NN) and average power. The power results on Tianjic were evaluated at the chip level in run time. The power results on the Xavier NX and AGX Orin were obtained via NVIDIA tegrastats.

Statistical analysis

We assessed the performance of NeuroVPR using accuracy, precision-recall curves, and Recall@K measures, which are commonly used metrics in previous VPR studies (69). Each model was trained five times, and we report the average results along with the SDs in all figures. The error bars displayed in each figure

represent the SD across the five trials. Sample sizes are described in Results and in the figure legends.

Supplementary Materials

This PDF file includes:

Tables S1 and S2
Figs. S1 to S10

REFERENCES AND NOTES

1. S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, M. J. Milford, Visual place recognition: A survey. *IEEE Trans. Robot.* **32**, 1–19 (2016).
2. T. Fischer, S. Garg, M. Milford, Where is your place, visual place recognition, in International Joint Conference on Artificial Intelligence (IJCAI), 19 to 26 August 2021, pp. 4416–4425.
3. L. Wu, T. Wang, C. Sun, Multimodal visual place recognition in dynamics-invariant perception space. *IEEE Signal Process. Lett.* **28**, 2197–2201 (2021).
4. M. Xu, T. Fischer, N. Sünderhauf, M. Milford, Probabilistic appearance-invariant topometric localization with new place awareness. *IEEE Robot. Autom. Lett.* **6**, 6985–6992 (2021).
5. P. Yin, S. Zhao, I. Cisneros, A. Abuduweili, G. Huang, M. Milford, C. Liu, H. Choset, S. Scherer, General place recognition survey: Towards the real-world autonomy age. arXiv:2209.04497 (2022).
6. M. Milford, G. Wyeth, Persistent navigation and mapping using a biologically inspired slam system. *Int. J. Robot. Res.* **29**, 1131–1153 (2010).
7. M. Milford, G. Wyeth, SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights, in IEEE International Conference on Robotics and Automation (ICRA), Saint Paul, MN, USA, 14May to 18 May 2012, pp. 1643–1649.
8. F. Yu, J. Shang, Y. Hu, M. Milford, NeuroSLAM: A brain-inspired SLAM system for 3D environments. *Biol. Cybern.* **113**, 515–545 (2019).
9. M. Chancán, L. Hernandez-Nunez, A. Narendra, A. B. Barron, M. Milford, A hybrid compact neural architecture for visual place recognition. *IEEE Robot. Autom. Lett.* **5**, 993–1000 (2020).
10. M. Chancán, M. Milford, DeepSeqSLAM: A trainable CNN+RNN for joint global description and sequence-based place recognition, in the Thirty-fourth Conference on Neural Information Processing Systems (NeurIPS) Workshop on Machine Learning for Autonomous Driving (ML4AD), 6 to 12 December 2020, pp. 1–9.
11. M. Chancán, M. Milford, Sequential place learning: Heuristic-free high-performance long-term place recognition. arXiv:2103.02074 [cs.CV] (2021).
12. A. Özdemir, M. Scerri, A. B. Barron, A. Philippides, M. Mangan, E. Vasilaki, L. Manneschi, EchoVPR: Echo state networks for visual place recognition. *IEEE Robot. Autom. Lett.* **7**, 4520–4527 (2022).
13. G. Gallego, T. Delbruck, G. M. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. Davison, J. Conradt, K. Daniilidis, D. Scaramuzza, Event-based vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 154–180 (2022).
14. D. Falanga, K. Kleber, D. Scaramuzza, Dynamic obstacle avoidance for quadrotors with event cameras. *Sci. Robot.* **5**, eaaz9712 (2020).
15. X. Zou, T. Huang, S. Wu, Towards a new paradigm for brain-inspired computer vision. *Machine Intelligence Research* **19(5)**, 412–424 (2022).
16. M. Milford, H. Kim, M. Mangan, S. Leutenegger, T. Stone, B. Webb, A. Davison, Place recognition with event-based cameras and a neural implementation of SeqSLAM, arXiv:1505.04548 [cs.RO] (2015).
17. A. Jacobson, Z. Chen, M. Milford, Leveraging variable sensor spatial acuity with a homogeneous, multi-scale place recognition framework. *Biol. Cybern.* **112**, 209–225 (2018).
18. T. Fischer, M. Milford, Event-based visual place recognition with ensembles of temporal windows. *IEEE Robot. Autom. Lett.* **5**, 6924–6931 (2020).
19. A. J. Lee, A. Kim, EventVLAD: Visual place recognition with reconstructed edges from event cameras, in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September to 1 October 2021, pp. 2247–2252.
20. D. Kong, Z. Fang, K. Hou, H. Li, J. Jiang, S. Coleman, D. Kerr, Event-VPR: End-to-end weakly supervised deep network architecture for visual place recognition using event-based vision sensor. *IEEE Trans. Instrum. Meas.* **71**, 5011418 (2022).
21. T. Fischer, M. Milford, How many events do you need? Event-based visual place recognition using sparse but varying pixels. *IEEE Robot. Autom. Lett.* **7**, 12275–12282 (2022).
22. S. Hussaini, M. Milford, T. Fischer, Spiking neural networks for visual place recognition via weighted neuronal assignments. *IEEE Robot. Autom. Lett.* **7**, 4094–4101 (2022).
23. J. J. DiCarlo, D. Zoccolan, N. C. Rust, How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).

24. N. J. Killian, M. J. Jutras, E. A. Buffalo, A map of visual space in the primate entorhinal cortex. *Nature* **491**, 761–764 (2012).
25. P. S. B. Finnie, R. W. Komorowski, M. F. Bear, The spatiotemporal organization of experience dictates hippocampal involvement in primary visual cortical plasticity. *Curr. Biol.* **31**, 3996–4008.e6 (2021).
26. N. M. Dotson, M. M. Yartsev, Nonlocal spatiotemporal representation in the hippocampus of freely flying bats. *Science* **373**, 242–247 (2021).
27. P. Georges-François, E. T. Rolls, R. G. Robertson, Spatial view cells in the primate hippocampus: Allocentric view not head direction or eye position or place. *Cereb. Cortex* **9**, 197–212 (1999).
28. E. T. Rolls, S. M. Stringer, Spatial view cells in the hippocampus, and their idiothetic update based on place and head direction. *Neural Netw.* **18**, 1229–1241 (2005).
29. R. Fettiplace, C. M. Hackney, The sensory and motor roles of auditory hair cells. *Nat. Rev. Neurosci.* **7**, 19–29 (2006).
30. S. Danilovich, Y. Yovel, Integrating vision and echolocation for navigation and perception in bats. *Sci. Adv.* **5**, eaaw6503 (2019).
31. R. B. Sulser, B. D. Patterson, D. J. Urban, A. I. Neander, Z.-X. Luo, Evolution of inner ear neuroanatomy of bats and implications for echolocation. *Nature* **602**, 449–454 (2022).
32. X. Bao, E. Gjorgieva, L. K. Shanahan, J. D. Howard, T. Kahnt, J. A. Gottfried, Grid-like neural representations support olfactory navigation of a two-dimensional odor space. *Neuron* **102**, 1066–1075.e5 (2019).
33. Y. Wu, K. Chen, Y. Ye, T. Zhang, W. Zhou, Humans navigate with stereo olfaction. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 16065–16071 (2020).
34. C. Poo, G. Agarwal, N. Bonacchi, Z. F. Mainen, Spatial maps in piriform cortex during olfactory navigation. *Nature* **601**, 595–599 (2022).
35. J. O’Keefe, J. Dostrovsky, The hippocampus as a spatial map, preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* **34**, 171–175 (1971).
36. M. M. Yartsev, N. Ulanovsky, Representation of three-dimensional space in the hippocampus of flying bats. *Science* **340**, 367–372 (2013).
37. J. S. Taube, R. U. Muller, J. B. Ranck Jr., Head-direction cells recorded from the postsubiculum in freely moving rats. I. description and quantitative analysis. *J. Neurosci.* **10**, 420–435 (1990).
38. A. Finkelstein, D. Derdikman, A. Rubín, J. N. Foerster, L. Las, N. Ulanovsky, Three-dimensional head-direction coding in the bat brain. *Nature* **517**, 159–164 (2015).
39. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
40. G. Ginosar, J. Aljadeff, Y. Burak, H. Sompolinsky, L. Las, N. Ulanovsky, Locally ordered representation of 3D space in the entorhinal cortex. *Nature* **596**, 404–409 (2021).
41. A. Tsao, J. Sugar, L. Lu, C. Wang, J. J. Knierim, M.-B. Moser, E. I. Moser, Integrating time from experience in the lateral entorhinal cortex. *Nature* **561**, 57–62 (2018).
42. J. B. Issa, G. Tocker, M. E. Hasselmo, J. G. Heys, D. A. Dombeck, Navigating through time: A spatial navigation perspective on how the brain may encode time. *Annu. Rev. Neurosci.* **43**, 73–93 (2020).
43. D. B. Omer, L. Las, N. Ulanovsky, Contextual and pure time coding for self and other in the hippocampus. *Nat. Neurosci.* **26**, 285–294 (2022).
44. A. A. Kinkhabwala, Y. Gu, D. Aronov, D. W. Tank, Visual cue-related activity of cells in the medial entorhinal cortex during navigation in virtual reality. *eLife* **9**, e43140 (2020).
45. S. Inayat, B. B. McAllister, B. L. McNaughton, I. Q. Whishaw, M. H. Mohajerani, Conjunctive and complementary CA1 hippocampal cell populations relate sensory events to immobility and locomotion. *bioRxiv* 2022.07.06.498996 (2022).
46. J. Pei, L. Deng, S. Song, M. Zhao, Y. Zhang, S. Wu, G. Wang, Z. Zou, Z. Wu, W. He, F. Chen, N. Deng, S. Wu, Y. Wang, Y. Wu, Z. Yang, C. Ma, G. Li, W. Han, H. Li, H. Wu, R. Zhao, Y. Xie, L. Shi, Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* **572**, 106–111 (2019).
47. L. Deng, G. Wang, G. Li, S. Li, L. Liang, M. Zhu, Y. Wu, Z. Yang, Z. Zou, J. Pei, Z. Wu, X. Hu, Y. Ding, W. He, Y. Xie, L. Shi, Tianjic: A unified and scalable chip bridging spike-based and continuous neural computation. *IEEE J. Solid-State Circuits* **55**, 2228–2246 (2020).
48. G. Wang, S. Ma, Y. Wu, J. Pei, R. Zhao, L. Shi, End-to-end implementation of various hybrid neural networks on a cross-paradigm neuromorphic chip. *Front. Neurosci.* **15**, 615279 (2021).
49. M. Davies, A. Wild, G. Orchard, Y. Sandamirskaya, G. A. F. Guerra, P. Joshi, P. Plank, S. R. Risbud, Advancing neuromorphic computing with Loihi: A survey of results and outlook. *Proc. IEEE* **109**, 911–934 (2021).
50. Y. Sandamirskaya, M. Kaboli, J. Conradt, T. Celikel, Neuromorphic computing hardware and neural architectures for robotics. *Sci. Robot.* **7**, eab18419 (2022).
51. S. Ma, J. Pei, W. Zhang, G. Wang, D. Feng, F. Yu, C. Song, H. Qu, C. Ma, M. Lu, F. Liu, W. Zhou, Y. Wu, Y. Lin, H. Li, T. Wang, J. Song, X. Liu, G. Li, R. Zhao, L. Shi, Neuromorphic computing chip with spatiotemporal elasticity for multi-intelligent-tasking robots. *Sci. Robot.* **7**, eabk2948 (2022).
52. C. Bartolozzi, G. Indiveri, E. Donati, Embodied neuromorphic intelligence. *Nat. Commun.* **13**, 1024 (2022).
53. R. Kreiser, A. Renner, Y. Sandamirskaya, P. Pienroj, Pose estimation and map formation with spiking neural networks: Towards neuromorphic SLAM, in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1 to 5 October 2018, pp. 2159–2166.
54. S. Baumgartner, A. Renner, R. Kreiser, D. Liang, G. Indiveri, Y. Sandamirskaya, Visual pattern recognition with on-chip learning: Towards a fully neuromorphic approach, in 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Seville, Spain, 12 to 14 October 2020, pp. 1–5.
55. E. P. Frady, G. Orchard, D. Florey, N. Imam, R. Liu, J. Mishra, J. Tse, A. Wild, F. T. Sommer, M. Davies, Neuromorphic nearest neighbor search using Intel’s Pohoiki Springs, in Proceedings of the Neuro-inspired Computational Elements Workshop (NICE’20), Heidelberg Germany, 17 to 20 March 2020, **23**, pp. 1–10.
56. N. Imam, T. A. Cleland, Rapid online learning and robust recall in a neuromorphic olfactory circuit. *Nat. Mach. Intell.* **2**, 181–191 (2020).
57. G. Tang, N. Kumar, K. P. Michmizos, Reinforcement co-learning of deep and spiking neural networks for energy-efficient mapless navigation with neuromorphic hardware, in IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2020, pp. 6090–6097.
58. A. Vitale, A. Renner, C. Nauer, D. Scaramuzza, Y. Sandamirskaya, Event-driven vision and control for UAVs on a neuromorphic chip, in IEEE International Conference on Robotics and Automation, Xi’an, China, 30 May to 5 June 2021, pp. 103–109.
59. M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, A. Ng, ROS: An open-source Robot Operating System, in IEEE International Conference on Robotics and Automation (ICRA) Workshop on Open Source Software, Kobe, Japan, 12 to 17 May, 2009, **3(3.2)**, pp. 1–5.
60. H. Qiao, Y. Wu, S. Zhong, P. Yin, J. Chen, Brain-inspired intelligent robotics: Theoretical analysis and systematic application. *Machine Intelligence Research* **20(1)**, 1–18 (2023).
61. Y. Wu, L. Deng, G. Li, J. Zhu, L. Shi, Spatio-Temporal backpropagation for training high-performance spiking neural networks. *Front. Neurosci.* **12**, 331 (2018).
62. Y. Wu, L. Deng, G. Li, J. Zhu, Y. Xie, L. Shi, Direct training for spiking neural networks: Faster, larger, better, in Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA, 27 January to 1 February 2019, **33(01)**, pp. 1311–1318.
63. S. Wu, K. Y. M. Wong, C. C. A. Fung, Y. Mi, W. Zhang, Continuous attractor neural networks: Candidate of a canonical model for neural information representation. *F1000Research*, **5** (2016).
64. C. Teeter, R. Iyer, V. Menon, N. Gouwens, D. Feng, J. Berg, A. Szafer, N. Cain, H. Zeng, M. Hawrylycz, C. Koch, S. Mihalas, Generalized leaky integrate-and-fire models classify multiple neuron types. *Nat. Commun.* **9**, 709 (2018).
65. P. U. Diehl, M. Cook, Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front. Comput. Neurosci.* **9**, 99 (2015).
66. R. K. Srivastava, J. Masci, S. Kazerounian, F. Gomez, J. Schmidhuber, Compete to compute, in NIPS’13: Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, Nevada, 5 to 10 December 2013, **2**, pp. 2310–2318.
67. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980 (2014).
68. H. Li, S. Ma, T. Wang, W. Zhang, G. Wang, C. Song, HASP: Hierarchical asynchronous parallelism for multi-NN tasks. *TechRxiv* (2023); <https://doi.org/10.36227/techrxiv.22338757.v1>.
69. M. Zaffar, S. Garg, M. Milford, J. Kooij, D. Flynn, K. McDonald-Maier, S. Ehsan, VPR-Bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change. *Int. J. Comput. Vis.* **129**, 2136–2174 (2021).

Acknowledgments

Funding: This work was supported by the National Nature Science Foundation of China (nos. 62088102, 61836004, and 42201456), the National Key Research and Development Program of China (no. 2021ZD0200300), the CETC Haikang Group-Brain Inspired Computing Joint Research Center, and the China Postdoctoral Science Foundation (no. 2021 M701835). **Author contributions:** F.Y., Y.W., R.Z., and L.S. designed the NeuroGPR system architecture and the MHNN. Y.W. implemented the MHNN and algorithms. S.M., H.L., H.Q., C.S., and T.W. implemented the MHNN on Tianjic. F.Y., S.M., and H.L. integrated the NeuroGPR system on the robot. F.Y., Y.W., S.M., H.L., and M.X. worked on the experiments. F.Y., Y.W., S.M., R.Z., and L.S. contributed to the analysis and interpretation of results. F.Y., Y.W., S.M., R.Z., and L.S. wrote the manuscript with input from all authors. L.S. and R.Z. proposed the concept of MHNNs and hybrid neuromorphic chip architecture and supervised the whole project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to support the conclusions of this manuscript are included in the main text or Supplementary Materials. The datasets are available on Zenodo: <https://doi.org/10.5281/zenodo.7825811>.

Submitted 5 October 2021
Resubmitted 21 January 2023
Accepted 14 April 2023

Published 10 May 2023
10.1126/scirobotics.abm6996

Downloaded from <https://www.science.org> at The Hong Kong University of Science and Technology (Guangzhou) on May 25, 2026

Brain-inspired multimodal hybrid neural network for robot place recognition

Fangwen Yu, Yujie Wu, Songchen Ma, Mingkun Xu, Hongyi Li, Huanyu Qu, Chenhang Song, Taoyi Wang, Rong Zhao, and Luping Shi

Sci. Robot. **8** (78), eabm6996. DOI: 10.1126/scirobotics.abm6996

View the article online

<https://www.science.org/doi/10.1126/scirobotics.abm6996>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science Robotics (ISSN 2470-9476) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Robotics* is a registered trademark of AAAS.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works