nature communications

Article

https://doi.org/10.1038/s41467-024-51210-2

Terahertz spoof plasmonic neural network for diffractive information recognition and processing

Received: 18 December 2023

Accepted: 31 July 2024

Published online: 06 August 2024

Check for updates

Xinxin Gao[1,2,3], Ze Gu[2,3], Qian Ma[2,3] ✉, Bao Jie Chen[1], Kam-Man Shum[1], Wen Yi Cui[2], Jian Wei You[2], Tie Jun Cui[2] ✉ & Chi Hou Chan[1] ✉

All-optical diffractive neural networks, as analog artificial intelligence accelerators, leverage parallelism and analog computation for complex data processing. However, their low space transmission efficiency or large spatial dimensions hinder miniaturization and broader application. Here, we propose a terahertz spoof plasmonic neural network on a planar diffractive platform for direct multi-target recognition. Our approach employs a spoof surface plasmon polariton coupler array to construct a diffractive network layer, resulting in a compact, efficient, and easily integrable architecture. We designed three schemes: basis vector classification, multi-user recognition, and MNIST handwritten digit classification. Experimental results reveal that the terahertz spoof plasmonic neural network successfully classifies basis vectors, recognizes multi-user orientation information, and directly processes handwritten digits using a designed input framework comprising a metal grating array, transmitters, and receivers. This work broadens the application of terahertz plasmonic metamaterials, paving the way for terahertz on-chip integration, intelligent communication, and advanced computing systems.

Artificial neural networks (ANNs), driven by the development of artificial intelligence, have been explored for speech recognition[1], image sensing[2], and computer vision[3]. For traditional digital computers, power consumption will become a severe problem due to the introduction of many transistors[4]. Moreover, the chip manufacturing process will circumscribe the computers' scalability. Various alternative approaches to conventional digital hardware have been investigated to implement ANNs, encompassing diffractive neural networks and optical circuit neural networks[5–18], to handle these problems. All-optical diffractive deep neural networks, consisting of multiple layers of diffractive surfaces, have been reported to perform various functions at the speed of light and low power consumption[6]. More importantly, the diffractive network enables parallel information processing, significantly enhancing computational efficiency and speed. Integrating active chips into such networks gives rise to programmable diffractive systems based on digital-coding metasurface array[19], effectively manipulating and sensing electromagnetic waves. After that, such a diffractive network was widely explored to expand more applications[13,19–21]. Meanwhile, the misalignment issue encountered by three-dimensional diffractive neural networks could be mitigated by adopting two-dimensional architectures[22,23]. Nonetheless, these architectures exhibit a relatively low utilization rate of radiation energy and are also susceptible to interference with external devices.

Concurrently, optical neural networks based on planar circuits have been explored for chip-integrated designs, featuring Mach-Zehnder interferometers as their fundamental building blocks[16,24,25]. The physical structure of this network necessitates a specific design to facilitate matrix multiplication based on singular value decomposition[16]. The above operation will limit the scalability of the large-scale network[26]. Additionally, for a planar neural network tasked

[1]State Key Laboratory of Terahertz and Millimeter Waves, City University of Hong Kong, Hong Kong, China. [2]State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China. [3]These authors contributed equally: Xinxin Gao, Ze Gu, Qian Ma. ✉e-mail: maqian@seu.edu.cn; tjcui@seu.edu.cn; eechic@cityu.edu.hk

with classifying the handwritten digits, a preprocessing step in a traditional computer is typically used to transform the digit information into the input phase and amplitude information[10,23,26].

As a powerful candidate in compact systems, spoof surface plasmon polariton (SSPP) devices emerge, which can emulate the dispersion behaviors of natural optical surface plasmon polaritons (SPPs)[27] at the microwave and terahertz (THz) frequencies[27–29]. SSPPs exhibit strong field confinement and can be manipulated by constructing different metal patterns[30]. Significantly, SSPPs offer reduced metal losses compared to their SPP counterparts, which suffer from substantial metal loss. The SSPP devices have been shown to provide distinct compactness compared to traditional microstrip structures[31,32]. Additionally, the SSPP structures' flexible dispersion behaviors and easy integration have facilitated the creation of programmable spoof plasmonic neural networks (SPNNs), which have programmable weight and nonlinear activation functions to process and detect microwave signals[33]. Additionally, SSPP waveguides have demonstrated the ability to transmit signals in parallel with minimal crosstalk, highlighting their potential for next-generation wireless and on-chip THz communications[34]. The THz spectrum offers relatively high bandwidth, facilitating rapid data transmission and processing[35,36]. This capability significantly enhances the neural networks by enabling faster training and inference times compared to lower frequency ranges. Consequently, SSPP structures hold the potential to serve as robust interconnections within large-scale THz ANNs.

In this work, we propose a THz SPNN, which comprises multiple hidden layers, incorporating SSPP diffractive layers and phase shift layers, enabling the processing of THz waves on a planar platform. Each diffractive layer comprises cascaded compact SSPP coupler arrays and has high transmission efficiency to imitate the full-connected architecture. Furthermore, the network exhibits scalability through the expansion of coupler numbers both horizontally and vertically. The proposed SPNN is beneficial for integrating THz components into compact and planar devices. By training the network, we derive the desired phase distributions inside the phase modulation layers and construct the THz-SPNN. Experimental results show that a five-layer fully connected SPNN can effectively perform vector classification tasks at 250–280 GHz frequency range. We also experimentally demonstrate the network capability for in-situ information processing of different radiation targets in a wireless transceiver framework. In particular, we fabricated and experimentally demonstrated a designed input framework for image data, such as handwritten digits. This consists of a metal grating array, transmitters, receivers, and the SPNN, capable of directly performing classification tasks for handwritten digits.

Results

The fundamental architecture of THz-SPNN

Benefiting from the electromagnetic manipulation capabilities and easy integration of the SSPP structures, we propose a fully connected planar THz-SPNN comprising multiple hidden layers, incorporating both SSPP coupler arrays (acting as diffractive layers) and phase shift layers, as illustrated in Fig. 1a. To imitate the diffractive behaviors in a homogenous medium layer (usually in the air or uniform substrate), we use a series of cascaded SSPP couplers to achieve full-connected weight distributions. Notably, SSPP couplers exhibit superior compactness compared to traditional microstrip couplers due to their strong field confinement. Meanwhile, the diffractive layer based on the SSPP coupler array achieves higher transmission efficiency than in air or other mediums (see Supplementary Note 1). This substantial increase in efficiency contributes to further reductions in power consumption and enhances measuring sensitivity, improving classification accuracy. Phase shift structures of varying heights are sequentially placed between the diffractive layers to ensure that the pre-trained parameters can be accurately mapped to physical structures. The fabrication of the THz SPNN is then carried out using a photolithographic method. Experimental results demonstrate that the proposed THz SPNN can successfully perform multiple target recognition and processing by constructing different input configurations. 1) Basis vectors classification. We employ distinct input structures to mimic four categories of basis vectors, and each category is mapped to its respective channel, operating at a broad frequency range. 2) Multiple users' recognition. The users' orientation information is captured through the receiver. This received information is then efficiently processed by the THz SPNN in the context of wireless communication tasks, thus facilitating user recognition and differentiation. 3) Handwritten digit patterns' classification. Handwritten digit patterns consisting of a metal grating array are integrated into a diffractive platform consisting of transmitters, collectors, and the THz SPNN. Unlike traditional recognition methods that rely on extensive preprocessing steps, our proposed diffractive framework directly processes the handwritten digit information. Due to the varying scattering effect of each digit, the collector receives distinct electromagnetic wave information, which is subsequently processed by the THz SPNN. This approach leverages the inherent differences in the scattering properties of each digit, enabling more intuitive classification.

The fundamental SSPP unit of the THz SPNN is shown in Fig. 1b, where the SSPP unit with an aluminum metal ground is depicted, featuring structural parameters such as $p = 25\,\mu m$, $w = 10\,\mu m$, and $d = 32\,\mu m$. It is etched onto the benzocyclobutene (BCB) polymer with a dielectric constant of 2.65 (details Supplementary Note 2 for more details). The compact nature of the SSPP structure is evident from its dispersion behaviors. As the groove depth (h) reaches 51 μm, with increasing frequency, the dispersion curve gradually shifts away from the light line, indicating stronger field confinement, as shown in Fig. 1c. When the groove depth h decreases to 0 μm, the SSPP characteristic vanishes, giving way to traditional surface waves represented by microstrips. Consequently, in comparison to microstrip ($h = 0\,\mu m$), the SSPP structure is expected to be more compact in size[31]. As a comparison, we designed the SSPP coupler and the traditional microstrip coupler, as illustrated in Fig. 1d and Supplementary Fig. S2c, with their structural parameters listed in Table S2. When the SSPP coupler is excited at port 1, transmission signals are acquired at other ports, as displayed in Fig. 1e. The transmission parameters of S_{21} and S_{41} are nearly identical, with a phase difference of approximately 90° at around 265 GHz. This coupler exhibits sound isolation (S_{21}) and low reflection (S_{11}). The SSPP coupler can achieve a reduction in size of about 11% when compared to its traditional counterparts (see Supplementary Note 2 for more details). Thus, the diffractive layer, comprising the SSPP coupler array, exhibits a reduced electric footprint and minimized interference with other devices. Then, we design a phase shift structure with a length (L_p) of 450 μm, enabling adjustments of the network's phase by varying the height (h_1), as shown in Fig. 1f. When h_1 varies from 40 to 150 μm, the phase difference can reach 390° and the amplitude only fluctuates by about 1.5 dB, as illustrated in Fig. 1g. Therefore, the phase shift layer shown in Fig. 1a can be tailored by constructing the phase shift structure with distinct height parameters. The fundamental SSPP coupler and phase shifter lay the foundation for building and measuring the SPNN. Before training the THz SPNN, we need to construct different inputs configurable to verify its capacity for information recognition and processing.

Information input configurations of three recognition tasks

1) Basis vector. Generally, the complex input data is represented by the fundamental basis vectors that capture vital features, which are then fed into neural networks to learn the necessary mappings or classification rules, as shown in Fig. 2a. This method reduces dimensionality,

https://doi.org/10.1038/s41467-024-51210-2

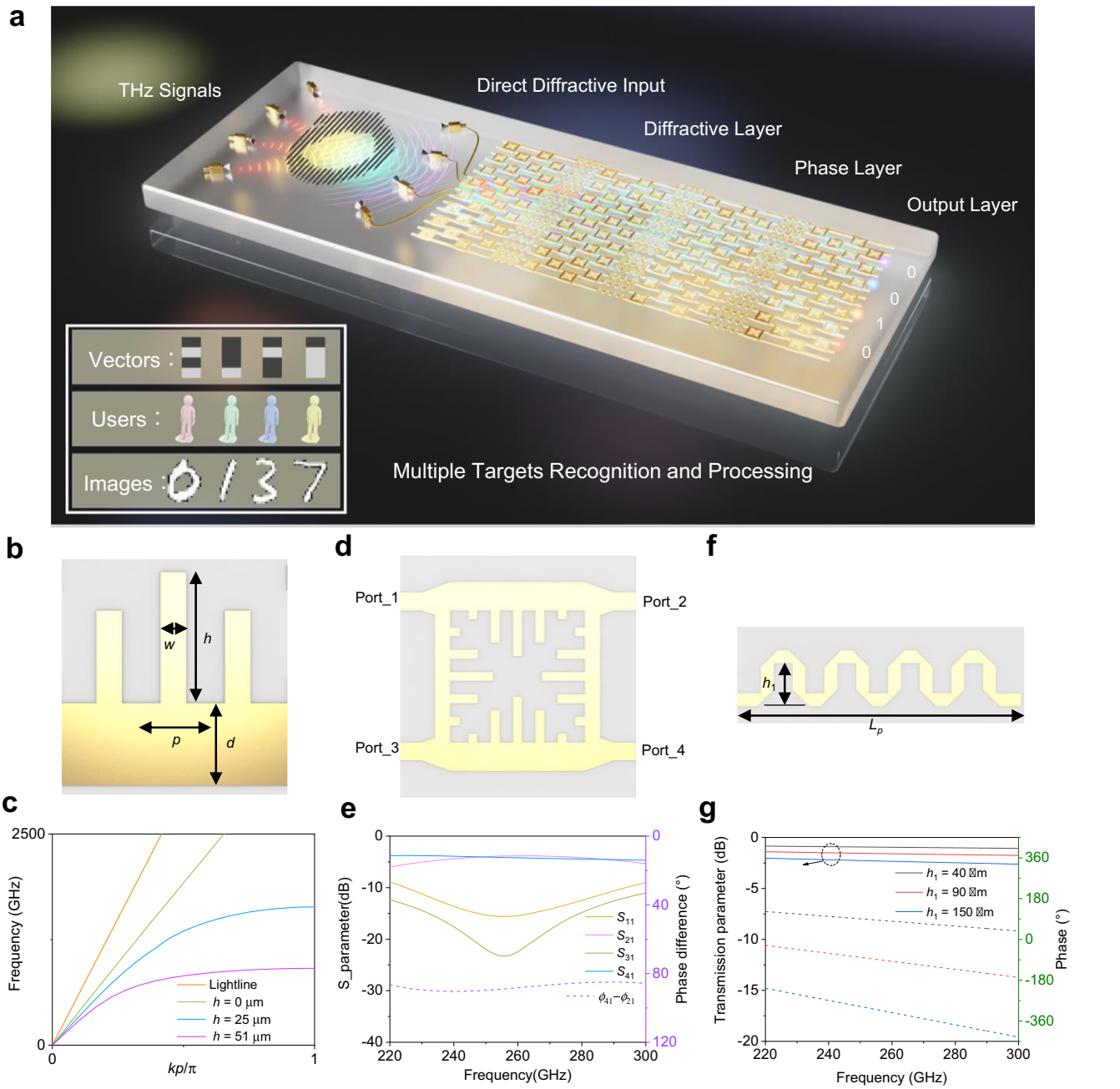

Fig. 1 | Conceptual illustration of the THz planar SPNN's multiple targets recognition and processing. a The integrated architecture is capable of performing classification tasks of basis vectors, multi-user recognition, and processing of handwritten digit images in the THz platform using the proposed SPNN. The integrated architecture mainly involves the transmitter, metal grating array, receiver, and the THz SPNN. The received electromagnetic waves are directly processed by the proposed THz-SPNN, consisting of one input layer, one output layer, and multi-hidden layers composed of diffractive and phase shift layers. **b** SSPP unit with structural parameters $p = 25\,\mu m$, $d = 32\,\mu m$, $w = 10\,\mu m$, and $h = 51\,\mu m$. **c** Dispersion behaviors of SSPP unit versus the groove depth h varying from 0 to 51 μm, where k and p are the wavenumber and period of the SSPP unit, respectively. **d** SSPP coupler. **e** Simulated S-parameter for the SSPP coupler. **f**, **g** Show the phase shift structure and the corresponding transmission and phase results, respectively, with varying heights h_1.

extracts essential features, and can be applied to various tasks efficiently. We employ four class basis vectors with approximate amplitude and distinct phase differences for the classification task. The phase differences of four categories are defined as follows: class 1 for [0°, 0°, 0°, 0°]; class 2 for [0°, 90°, 180°, 270°]; class 3 for [0°, 180°, 0°, 180°]; class 4 for [0°, −90°, −180°, −270°]. Due to the limitation of experimental conditions within the THz frequency, we use the various input structures to simulate basis vectors for four categories. As depicted in Fig. 2b, the input structure of class 1 is designed to incorporate cascaded microstrip couplers to ensure uniform output amplitude. Different phase shift structures are combined to introduce the designated phase difference. Moreover, to mitigate the reflection, the matched load with nickel metal is introduced (see Supplementary Note 3 for more details). The simulated results demonstrate an approximate transmission amplitude and phase around 265 GHz, indicating category 1, as illustrated in Fig. 2c. The input structures of the other three categories and their simulated results are detailed in Supplementary Note 4. The structural configuration of basis vectors provides a foundation for performing fundamental tasks in a THz platform.

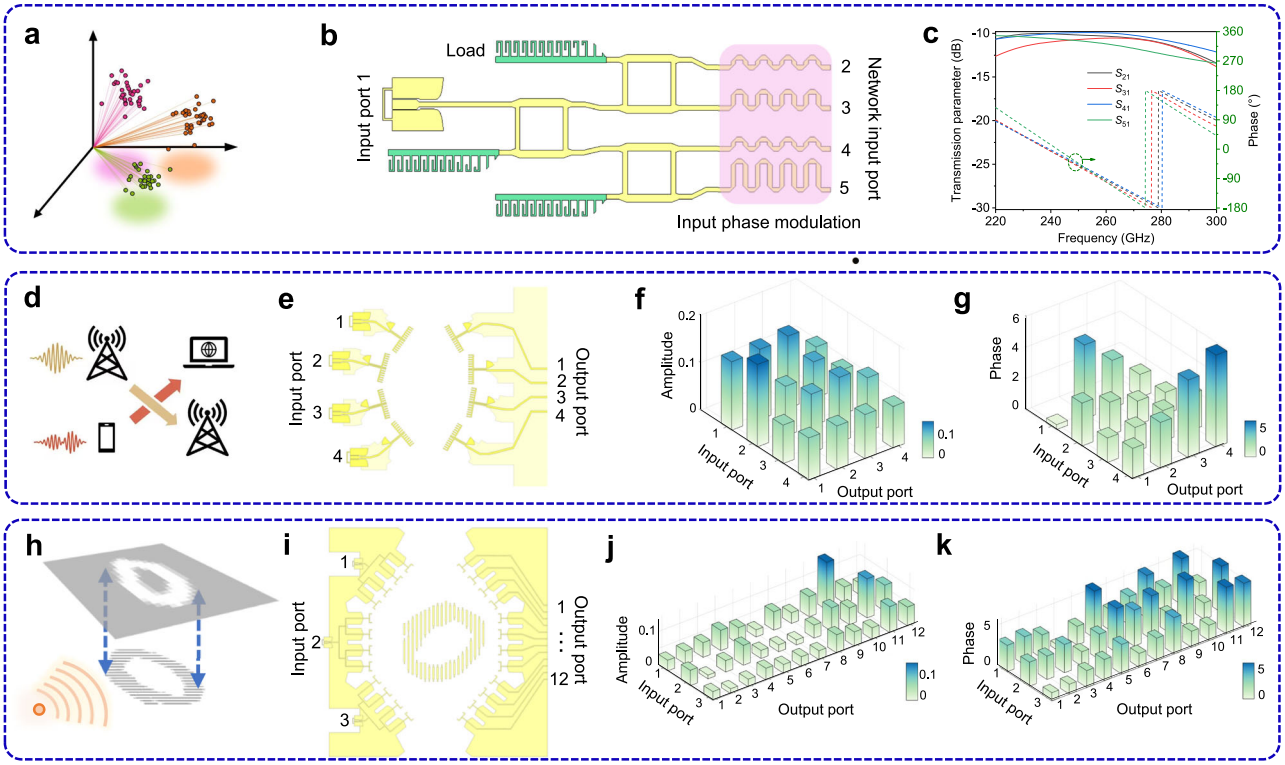

Fig. 2 | The different input structures to perform three recognition tasks. **a** Basis vector classification. **b** The input structure of class 1, in which network input ports are labeled as ports 2–5. **c** The transmission (S_{21}, S_{31}, S_{41}, and S_{51}) and phase results of the input structure, in which the blue structure represents the designed SSPP load featuring a nickel metal[42]. **d** Multi-user communication application. **e** The input structure to mimic the user information. **f**, **g** The amplitude and phase information between the input and output ports. **h** Handwritten digits' classification. **i** The diffractive architecture to directly recognize the image. **j**, **k** The collected amplitude and phase information by 12 receivers.

2) Multiple users' recognition. In wireless communications, accurate user-orientation recognition is usually of great significance to enhance communication channel quality and increase communication efficiency, as shown in Fig. 2d. In the traditional communication architecture[37,38], this process typically involves a series of measurements, sampling, and analog-to-digital conversion processes on the computer side to achieve orientation recognition. However, the entire process can be easily and swiftly implemented by deploying a simple diffraction neural network on the RF front end, thereby eliminating the need for more complex system hardware and processes, such as mixers and analog-to-digital converters. In this context, we design a compact 4-input and 4-output communication structure consisting of four transmitting users with different directions and four receiving antennas, as illustrated in Fig. 2e. The receiving antennas collect orientation information from the users, and the gathered amplitude and phase data (depicted in Figs 2f and g) are subsequently input into and directly processed by the SPNN.

3) Handwritten digits classification. In general, for a planar neural network to classify the handwritten digits, a preprocessing step in a traditional computer is typically used to transform the digit information (Fig. 2h) into the input phase and amplitude data[10,23,26]. Here, we present a diffractive architecture to process image data, such as MNIST handwritten digits, in a planar diffractive network, as shown in Fig. 2i, consisting of transmitters, recognition images consisting of a metal grating array, and 12 receiving antennas. To comprehensively collect the image information, three groups of antennas (each group using a 1/4 power divider excitation) are employed to direct the electromagnetic wave irradiation in various directions. The 12 receiving antennas then collect the amplitude and phase information of handwritten digits, as shown in Fig. 2j and k. It can be observed that the image data varies distinctly when different input ports are excited, indicating that diverse data information can be collected from the various excitation directions. For different digit patterns, their structural variations significantly influence the scattering electromagnetic field distributions, resulting in distinct amplitude and phase information (see Supplementary Note 9). This information can be directly processed by the SPNN. For each classification task, the input information for each category will be mapped to its respective output port through the direct processing of the THz SPNN.

To verify the above schemes, we established a comprehensive THz experimental setup and fabricated test samples via photolithography, as illustrated in Fig. 3. The setup entails the utilization of two GSG probes (as depicted in Fig. 3d), connected to two frequency extenders operating within a frequency range from 220 to 325 GHz. This configuration serves to generate and capture THz signals from the testing samples (Fig. 3b) by the vector network analyzer (VNA), as shown in Fig. 3c. One of the GSG probes is employed to excite the input port of the sample, allowing the electromagnetic signals to propagate along the input structure. The process generates diverse electromagnetic diffractive information. Following this, SPNN processes the electromagnetic signals and maps them to the corresponding channels. Finally, the other GSG probe is utilized to capture the output amplitude distribution at the output port. For the tasks involving basis vectors and user recognition, the diffractive layer comprises 4-cascaded SSPP coupler arrays. For more complex classification tasks, such as handwritten digits, the diffractive layer is composed of 6-cascaded SSPP coupler arrays.

Basis vector classification by the THz SPNN

To verify the capacity of the THz SPNN to classify basis vectors, we fabricate four 5-layer THz-SPNN samples, as shown in Fig. 4a and S6. Meanwhile, the loads with nickel are integrated into the corresponding

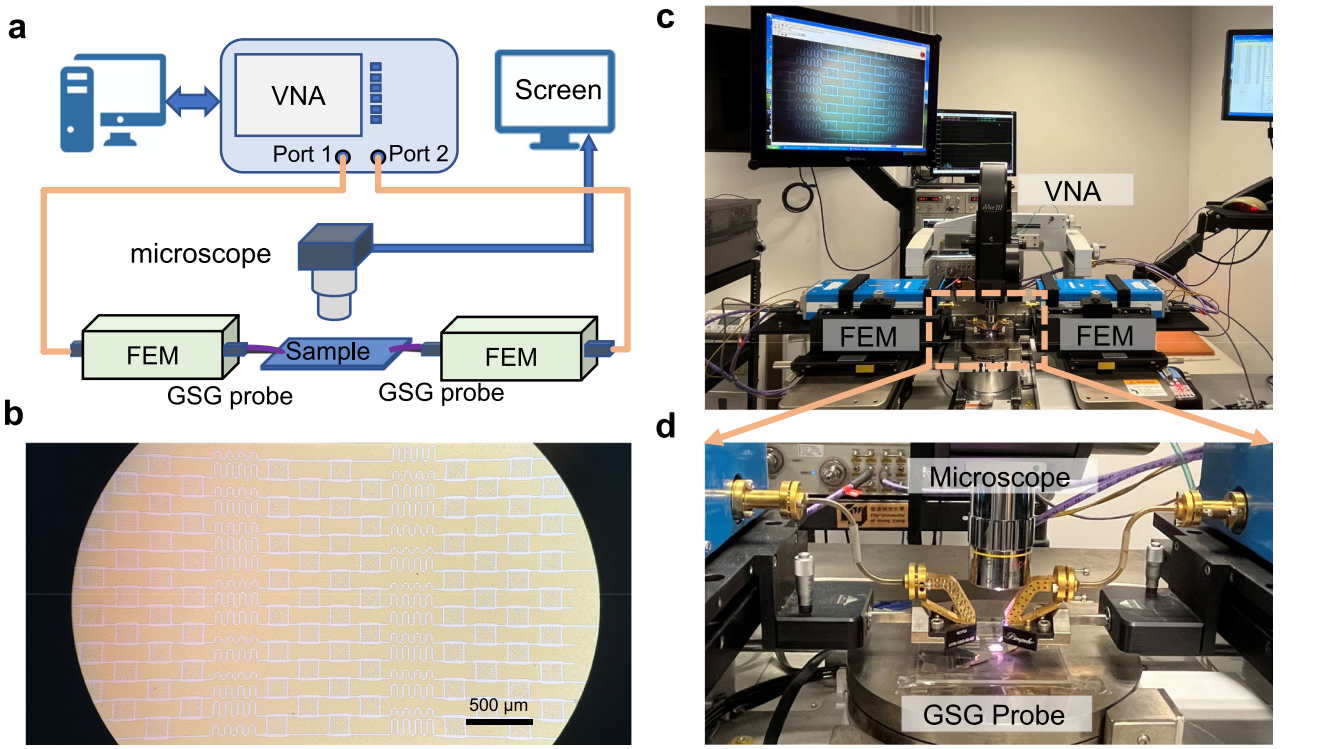

Fig. 3 | Experimental measurement scheme to obtain the classification performance of the THz-SPNN. a The schematic illustration of the measurement setup. The frequency extender module (FEM) can extend the working frequency of the vector network analyzer (VNA) to 220–325 GHz. Then, the transmission parameters of the THz SPNN can be obtained by using the ground-signal-ground (GSG) probes. **b** Partial sample display under the microscope. **c** Photograph of the experimental set-up. **d** GSG probe to excite the sample.

ports of the SPNN to reduce the reflection. We predict the capability of the proposed THz-SPNN through a customized vector classification task, where each category in a dataset is defined by adding a basis vector with Gaussian white noise. The proposed THz-SPNN is trained using the gradient-based backpropagation algorithm[39]. Note that this training is performed at a frequency of 265 GHz (see Methods for training details). The category accuracy of the SPNN with different layers is depicted in Fig. 4b. As the number of network layers increases from 3 to 5, the corresponding accuracy gradually improves, rising from 66% to 99%. We select a 5-layer network in which the classification reaches 99% accuracy, proving it is sufficient to accomplish the classification task. As expected, each category (class 1, class 2, class 3, and class 4) can be accurately mapped to its respective output channel (port 4, port 7, port 10, and port 13) and has a corresponding maximum amplitude distribution in each channel, as shown in Fig. 4c. The confusion matrix is calculated, and 99% accuracy can be achieved (Fig. 4d). To further verify the classification performance, we simulate the transmission amplitude distribution of the designed THz-SPNN, where each category is mapped to its respective output port within the frequency range of 250–280 GHz (see Supplementary Note 5). This suggests that the SPNN has the capability to perform classification tasks across a wide frequency band.

Subsequently, the transmission parameters of the THz-SPNN are measured, as presented in Fig. 4e–h. As expected, each category exhibits maximum amplitude distribution at the respective output port, operating at the frequency band of 250 to 280 GHz. While there are some disparities between measured and simulated results, likely due to errors in fabrication and measurement, overall, these results are in good agreement. Subsequently, we collect the amplitude distributions of the network at 250, 265, and 280 GHz, as illustrated in Figs. 4i–k. The majority of the energies for categories 1, 2, 3, and 4 are distributed to the output ports 4, 7, 10, and 13, respectively. The corresponding confusion matrices further validate the mapped relations, as depicted in Figs. 4l–n. At 250, 265, and 280 GHz, the classification accuracy can reach 95.8%, 99.9%, and 92.4%, respectively (see "Methods"). The classification performance of the SPNN is also demonstrated at 260 and 270 GHz (see Supplementary Note 6 for more details). Therefore, the proposed THz-SPNN successfully accomplishes classification tasks within a frequency range of 250–280 GHz.

In-situ information processing by the THz SPNN

We present an in-situ recognition scenario for different users' directions, including the transmitter, receiver, and SPNN, as shown in Fig. 5a. We show that the orientation information can be directly obtained on the output interface after the signals are transmitted through the networks without involvement of the additional devices. To simulate real scenarios, we design a transceiver array including four transmitting and four receiving SSPP antennas (see Supplementary Note 7). The four transmitting SSPP antennas are defined as users 1, 2, 3, and 4 with different orientations. The corresponding orientation information can be represented by collected amplitude and phase distributions using the receiving antennas. Then, the information from various users at the transmitter is efficiently detected and then processed by the SPNN. Each user's information is systematically mapped to its corresponding output channel. For instance, upon exciting user 1, its information is transmitted to four received SSPP antennas. Subsequently, the received information undergoes processing by the SPNN, allowing us to extract the user's information by analyzing the amplitude distribution across the four output ports. In this scenario, the maximum distribution is detected at output port 4. Similarly, users 2, 3, and 4 correspond to the output ports 7, 10, and 13, respectively. Measured results verify the inference mentioned above, as shown in Figs. 5b–e. We observe that the information processing capability of SPNN is not limited to a single frequency but can be extended across a broad frequency range. Discrepancies between measured and simulated insertion loss are analyzed (more details in Supplementary

https://doi.org/10.1038/s41467-024-51210-2

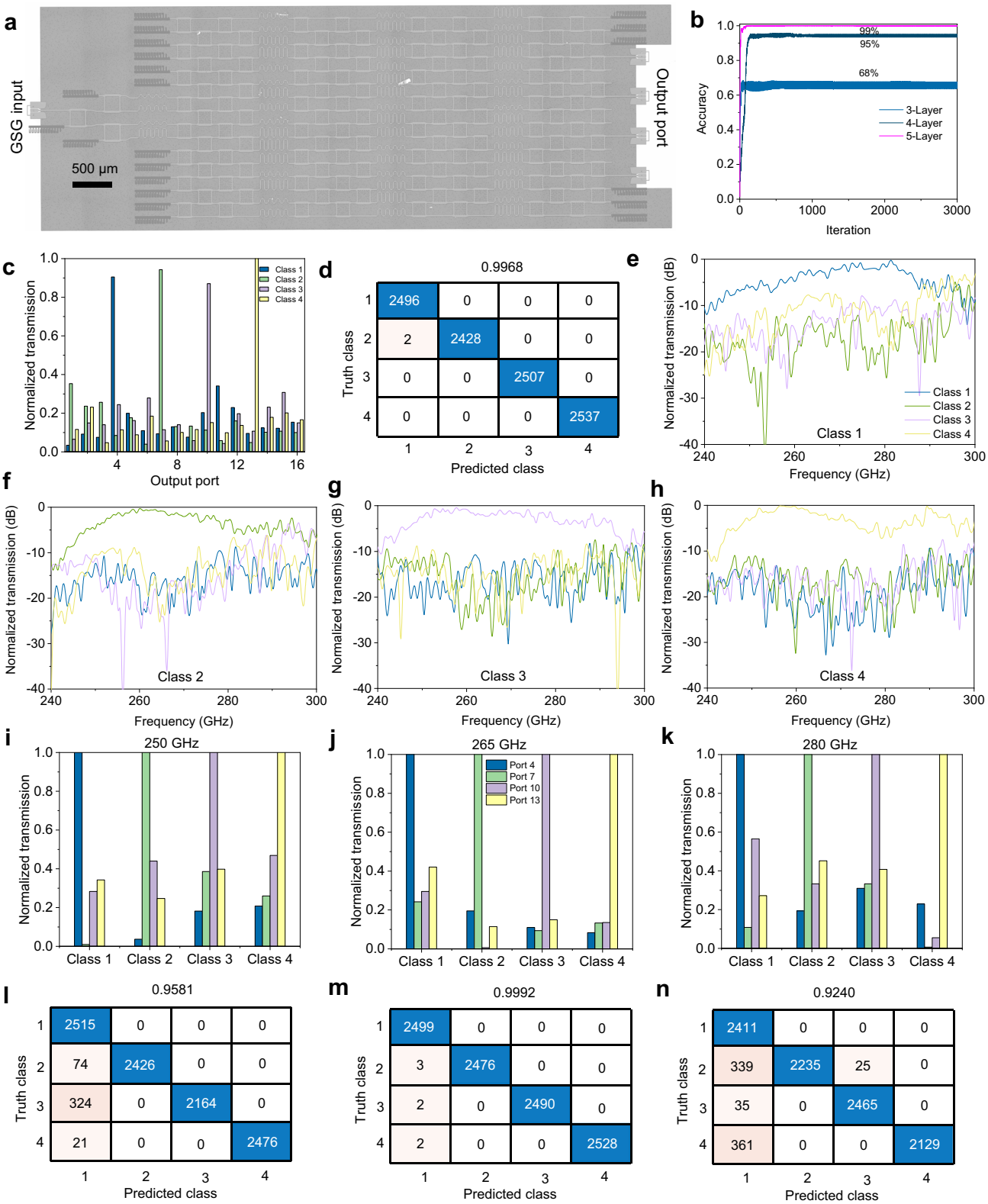

Fig. 4 | Experimental results. a The fabricated sample diagram of the 5-layer SPNN for category 1 under magnified perspective. **b** Calculated classification accuracy of different layers. **c** The output amplitude distributions at four classes. **d** The confusion matrix of the 5-layer SPNN. Since some misidentified ports are not in the 4 categories, the complete confusion matrix of all ports is given in Supplementary Fig. S5h. **e**–**h** Measurement of transmission parameters of four categories at output ports 4, 7, 10, and 13, respectively, in which each category has the maximum transmission parameters at its respective output port at a broadband frequency range. **i**–**k** are measured amplitude distributions at different categories when the operating frequencies are 250 GHz, 265 GHz, and 280 GHz, respectively. **l**–**n** are confusion matrices for 250 GHz, 265 GHz, and 280 GHz, respectively, corresponding to 95.8%, 99.9%, and 92.4% accuracy.

https://doi.org/10.1038/s41467-024-51210-2

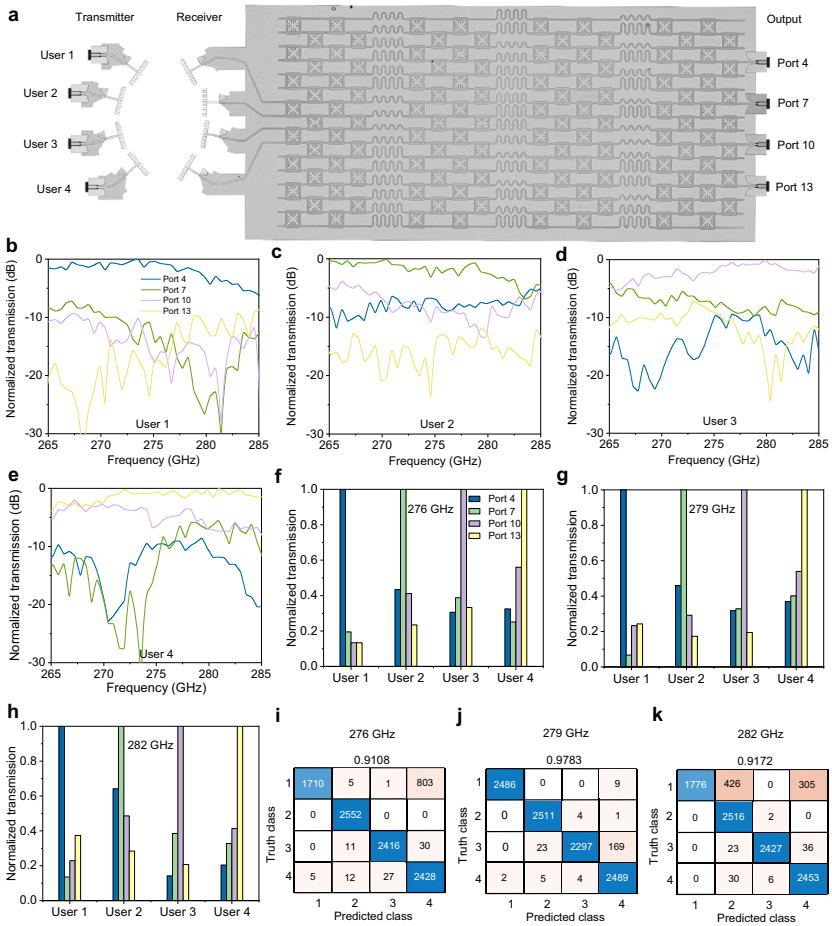

Fig. 5 | Experimental results of users' orientation recognition. a The fabrication sample, where the transmitting SSPP antennas (users 1, 2, 3, and 4) can radiate electromagnetic waves to the receiver consisting of four receiving SSPP antennas. **b**–**e** Measurement transmission parameters of four users at output ports 4, 7, 10, and 13, respectively. **f**–**h** Measured amplitude distributions of different users at 276 GHz, 279 GHz, and 282 GHz, respectively. **i**–**k** Confusion matrices at 276 GHz, 279 GHz, and 282 GHz, respectively, corresponding to classification accuracies of 91%, 97.8%, and 91.7%.

Note 7), primarily stemming from manufacturing and measurement technologies. Nonetheless, we can still identify a suitable frequency band to showcase the SPNN's capacity for information processing and recognition. As illustrated in Fig. 5f, it is evident that the amplitude distribution for each user has a peak at the respective port when the operating frequency is 276 GHz. Similarly, the amplitude distribution peaks occur at 279 GHz and 282 GHz, as depicted in Figs. 5g, h. This observation indicates the robustness of the THz-SPNN. Additionally, corresponding confusion matrices at these frequencies are calculated and presented in Figs. 5i–k, with achieved accuracies of 91%, 97.8%, and 91.7%, respectively (see Methods for more details). Consequently, SPNN successfully processes and recognizes user information.

It is noteworthy that the users at different orientations in the presented scheme are situated approximately 900 μm away from the receiving antenna arrays, primarily constrained by the sensitivity of the THz measuring system. The proposed network exhibits the potential to process signals at greater distances, as substantiated by numerical simulations (refer to Supplementary Note 8). In the realm of physical neural networks, the principal factors influencing recognition accuracy are signal distortion and noise interference.

Handwritten digits classification by the THz SPNN

Here, we present an input data method for image data, such as MNIST handwritten digits in a planar diffractive network, as depicted in Fig. 6a. This corresponding architecture comprises transmitters, recognition images consisting of a metal grating array, collectors, and an SPNN processing unit. Compared to a 5-layer network using classifying the basis vectors, this SPNN is scaled both horizontally and vertically. It has been demonstrated that a 7-layer SPNN can successfully classify the 10 handwritten digits with an accuracy of 96.6%, as shown in Fig. 6b, where its diffractive layer comprises a 6-cascaded SSPP coupler array. Then, we verify the classification capacity of the SPNN by the all-wave simulation, and each digit can be successfully mapped to its respective output port (see Supplementary Note 9). When the diffractive layer is reduced from 6 to 2-cascaded SSPP coupler arrays, the classification accuracy decreases by 50.5% due to network performance degradation (see Supplementary Fig. S20). This classification accuracy can be improved by increasing the number of network layers (see Supplementary Note 10).

However, due to THz experimental limitations (see Supplementary Notes 9 and 11), the SPNN was decreased to 5 layers and tasked with classifying four categories of digits ("0", "1", "3", and "7"), as shown in Fig. 6c. Through training, the network achieves an accuracy of 95% when configured with 5 layers, as shown in Fig. 6d and e (see "Methods" for training details). Each digit ("0", "1", "3", and "7") can be mapped to its respective output port (8, 11, 14, and 17) (see Supplementary Note 9). Subsequently, we fabricated and measured samples with different digits. Due to the varying scattering effect of each digit, the collector receives distinct electromagnetic diffractive information, which serves as inputs to the THz SPNN and is then processed and mapped to their respective channels, as depicted in Fig. 6f–i. While the measured results exhibit some discrepancies compared to the

https://doi.org/10.1038/s41467-024-51210-2

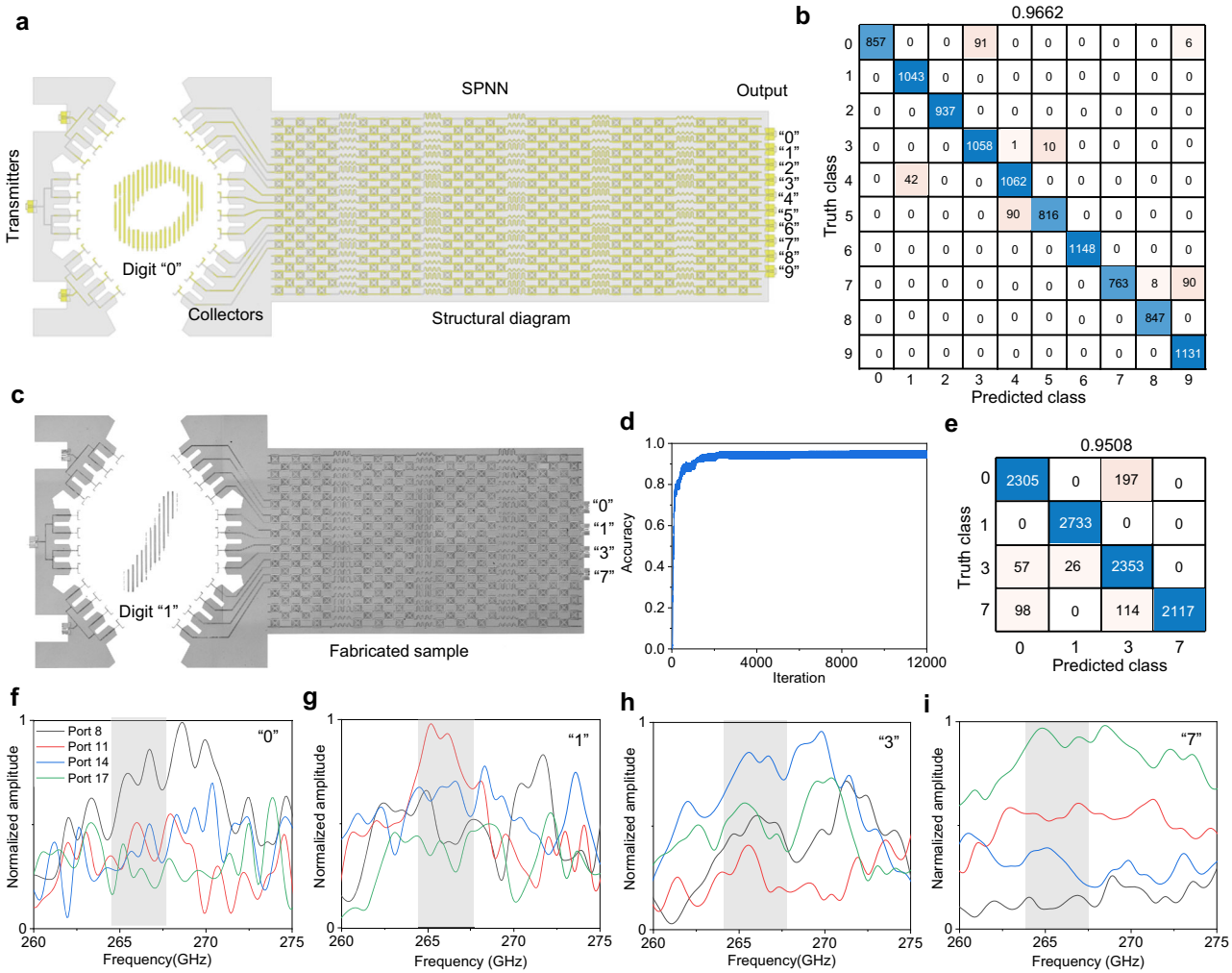

Fig. 6 | Experimental demonstration of MNIST handwritten digit recognition on an integrated diffractive architecture. a The structural diagram consisting of transmitters, recognition targets, collectors, and the processing information's SPNN. **b** The calculated confusion matrix when the 7-layer SPNN performs handwritten digits. **c** The fabrication sample for a 5-layer SPNN. **d**, **e** are corresponding accuracy and the confusion matrix when the THz SPNN performs the classification tasks of four digits ("0", "1", "3", and "7"). **f**–**i** are measurement transmission parameters of four digits at output ports 8, 11, 14, and 17, respectively, where the maximum energy for each digit is allocated to its corresponding port at a certain frequency band and the shading area indicates the operating frequency band.

simulated results (see Supplementary Fig. S16) due to manufacturing tolerance and measurement errors, they are acceptable for classification tasks performed by our proposed SPNN on the THz platform.

Discussion

We proposed an SPNN for multiple diffractive information recognizing processing in the THz frequencies. Leveraging the strong field confinements and flexible dispersion behaviors of SSPPs, we designed the SSPP coupler as the compact fundamental unit of the diffractive layer. Compared to the space transmission between the spatial diffractive structures, the surface transmission between the proposed diffractive layer has higher transmission efficiency. The THz SPNN can successfully perform different tasks, such as processing vectors and images directly from input information. When the input structures with different phases are introduced into the network, the SPNN can successfully classify the basis vectors at a wide frequency band of 250–280 GHz. Furthermore, we illustrate the SPNN's capability to recognize and process user information from the transmitting end in a wireless transceiver framework. In particular, we propose a diffractive architecture comprising transmitters, digit samples with a metal grating array, collectors, and the SPNN in a planar THz platform. The signals collected from the transmitter can be directly processed by the SPNN, showcasing its ability to classify the MNIST handwritten digits.

In comparison to optical MZI-based networks, the scalability of the SPNN is notably more straightforward. For example, the network size can be expanded by cascading SSPP couplers both horizontally and vertically. We note that a 5-layer SPNN with the diffractive layers consisting of four-cascaded SSPP coupler arrays has a 15 dB insertion loss, in which the network can be extended to approximately 13 layers at a −50 dB detector's sensitivity (see Supplementary Note 11 for details). Moreover, the input noise and analog computing noise influence the robustness and scalability of the THz SPNN. When classifying handwritten digits, the network's accuracy decreases slightly as the input noise increases from 0.02 to 0.12, but it remains above 90%, indicating significant robustness. However, since neural networks process the analog signals, the noise will accumulate with the increasing number of network layers, limiting the scalability. Despite the inevitability of the analog computing noise, its impact can be mitigated by improving the network's transmission efficiency layer (see Supplementary Note 12 for more details). The scalability can be further improved by improving the network's transmission efficiency, such as using a substrate with smaller loss (e.g., cyclic olefin copolymer (COC)[40] and quartz), a metal material with high conductivity (e.g.,

gold), and decreasing the length of the diffractive layer (details see Supplementary Note 11).

We highlight the feasibility of reconfigurable amplitude/phase and nonlinear control at the THz frequencies, addressing the need for dynamic signal modulation and enhancing the overall system flexibility and performance. This control can be accomplished using techniques such as two-dimensional materials and complementary metal oxide semiconductor (CMOS) manufacturing technology. The amplitude and phase modulations can be realized by changing the conductivity of Vanadium (IV) oxide (VO_2) loaded on the SSPP waveguide and changing the capacitance of two-dimensional electron gas (2DEG) materials integrated into the SSPP structure, respectively. Moreover, mixers and multipliers in the RF circuits allow for the up-convert of lower frequency signals into THz signals, facilitating the amplitude and phase modulation through variable gain amplifiers and phase shifters in CMOS manufacturing technology (see Supplementary Note 13 for details). In particular, nonlinear activation functions at THz frequencies are anticipated through the integration of metamaterial split ring resonators on doped GaAs films[41].

In the context of user recognition tasks performed by the THz SPNN, the communication range can be extended by replacing lower-gain antennas with higher-gain Vivaldi antennas or employing a phased array[37]. On the other hand, the short communication lengths in the THz range offer unique opportunities for on-chip and short-range communications, supporting high-density device connections and detailed surface scans of items (see Supplementary Note 14 for details). The THz SPNN enhances THz technology's performance and capabilities in these scenarios by providing advanced in-situ information processing.

Despite the high-speed computation and parallel processing capabilities of space-diffractive neural networks, the THz SPNN offers easier integration, higher transmission efficiency, and the capability for on-chip communication fabrication. In comparison to planar diffractive networks, the THz SPNN achieves not only higher transmission efficiency but also directly performs classification tasks for handwritten digits using a designed diffractive architecture. Considering the flexibility of the diffractive layer consisting of the SSPP coupler arrays, the proposed THz SPNN can also establish novel calculation models, such as convolutional neural networks (see Supplementary Note 15 for more details). Owing to the flexible dispersion behaviors, strong field confinement, and smaller metal loss than the optical SPPs[27], our work extends the application of plasmonic metamaterials into the THz frequency range. It holds the potential to pave the way for achieving fast and robust machine learning processes.

Methods

Training details

In the first experiment of basis-vector classification tasks, the dataset is generated by introducing Gaussian noise with a normalized intensity of 0.12 on the designated basis vector sets, constituting 20,000 datasets. We randomly selected 10,000 sets as the test set and the rest of the data as the training set. For each category, the number of training samples is around 2500. The classification performance on the testing dataset is nearly identical to the training dataset due to the addition of white noise (see Supplementary Note 12). The training process is carried out on the 10,000 samples, while the real experiment (depicted in Fig. 4) is validated on basis vectors physically constructed through couplers.

In the second experiment of information processing, the dataset is similarly extended from the basic vector sets by adding white noise signals. Unlike the previous one, the basic vector sets in the second experiment are generated from the transmission matrix between transmitting and receiving antennas through the full-wave simulations. The 10,000 training samples are fed into the network during the training process, and the conducted experiment is performed on the transmit-receive antenna, as shown in Fig. 5a.

In the third experiment of MNIST dataset classification via sensing mechanism, the training and testing samples are all generated through full-wave simulation in the first place. The received energy at the corresponding ports is extracted through the CST simulation software, and 100 samples are simulated. The simulated samples are extended to 10,000 training samples and 10,000 testing samples, again, using the white noise addition, with a normalized intensity of 0.02. The accuracy on the training dataset is 94.77% (Supplementary Fig. S25), and the validation accuracy is 95.08% (Fig. 6d). We fabricated four samples, 1 for each category, with a digit pattern engraved on the substrate. The experiment results are demonstrated in Fig. 6f–i.

Sample fabrication

We employ the photolithographic method to fabricate the testing samples, and the detailed process is as follows. First, a ground pattern with a thickness of 0.3 μm composed of aluminum was deposited onto the quartz substrate. Next, BCB polymer with a thickness of 7 μm was spin-coasted onto the aluminum film and subsequently cured at 270° for a duration of 2 h. Finally, we used the same photolithography process to deposit the SPNN and input structure patterns with aluminum films, as well as SSPP load patterns with nickel film onto the BCB polymer.

Measurement method

To evaluate the classification capabilities of the proposed SPNN operating within the 250–280 GHz frequency range, we utilize the frequency extender (V03VNA2-T/R). The frequency extender expands the VNA (Agilent Technologies, N5245A), enabling us to conduct millimeter-wave S-parameter measurements in the 220–325 GHz band. The two frequency extension modules connect to two ground-signal-ground (GSG, 325B-GSG-50-BT) probes and leverage the inherent performance and features of the microwave network analyzer to measure the transmission parameters of the testing samples. Moreover, for the multi-user recognition task, different transmission responses can be obtained by exciting the four users' ports, respectively. In this process, we use the VNA to measure the users' orientation information. Finally, to illustrate the classification tasks involving four handwritten digits, the proposed integrated diffractive architecture incorporates three excited ports and four output ports, as depicted in Fig. 6c. Each port excites four antennas through a one-to-four power divider. When a GSG probe excites port 1, another GSG probe is connected to an output port of the sample and manually shifted from the first to fourth output ports. The same procedure is repeated for other instances. Upon completion of the measurement, the results from the three excitation ports are processed. Finally, the measurement results are displayed in Fig. 6f–i. Therefore, we conduct measurements on four samples, considering the complexity and limitations of the measurement conditions. Nonetheless, the measurement results also demonstrate the THz SPNN's capacity for processing and recognizing information.

References

1. Hinton, G. et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process. Mag.* **29**, 82–97 (2012).
2. Wang, T. et al. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photonics* **17**, 408–415 (2023).
3. Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).

4. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
5. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **15**, 102–114 (2021).
6. Xing, L., et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
7. Jingxi, L., et al. Spectrally encoded single-pixel machine vision using diffractive networks. *Sci. Adv.* **7**, eabd7690 (2021).
8. Rahman, M. S. S. & Ozcan, A. Time-lapse image classification using a diffractive neural network. *Adv. Intell. Syst.* **5**, https://doi.org/10.1002/aisy.202200387 (2023).
9. Bai, B. et al. All-optical image classification through unknown random diffusers using a single-pixel diffractive network. *Light Sci. Appl.* **12**, 69 (2023).
10. Williamson, I. A. D. et al. Reprogrammable electro-optic nonlinear activation functions for optical neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 1–12 (2020).
11. Pai, S. et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* **380**, 398–404 (2023).
12. Zheng, Z. et al. Dual adaptive training of photonic neural networks. *Nat. Mach. Intell.* https://doi.org/10.1038/s42256-023-00723-4 (2023).
13. Qian, C. et al. Dynamic recognition and mirage using neuro-metamaterials. *Nat. Commun.* **13**, 2694 (2022).
14. Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
15. Hughes, T. W., Minkov, M., Shi, Y. & Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **5**, https://doi.org/10.1364/optica.5.000864 (2018).
16. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
17. Ma, W. et al. Deep learning for the design of photonic structures. *Nat. Photonics* **15**, 77–90 (2020).
18. Chen, Y. et al. Photonic unsupervised learning variational autoencoder for high-throughput and low-latency image transmission. *Sci. Adv.* **9**, https://doi.org/10.1126/sciadv.adf8437 (2023).
19. Liu, C. et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electron.* **5**, 113–122 (2022).
20. Yan, T. et al. Fourier-space diffractive deep neural network. *Phys. Rev. Lett.* **123**, 023901 (2019).
21. Chen, Y. et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* **623**, 48–57 (2023).
22. Wang, Z., Chang, L., Wang, F., Li, T. & Gu, T. Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nat. Commun.* **13**, 2131 (2022).
23. Fu, T. et al. Photonic machine learning with on-chip diffractive optics. *Nat. Commun.* **14**, 70 (2023).
24. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
25. Bogaerts, W. et al. Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
26. Zhu, H. H. et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044 (2022).
27. Barnes, W. L., Dereux, A. & Ebbesen, T. W. Surface plasmon subwavelength optics. *Nature* **424**, 824–830 (2003).
28. Pendry, J. B., Martín-Moreno, L. & Garcia-Vidal, F. J. Mimicking surface plasmons with structured surfaces. *Science* **305**, 847–848 (2004).
29. Shen, X., Cui, T. J., Martin-Cano, D. & Garcia-Vidal, F. J. Conformal surface plasmons propagating on ultrathin and flexible films. *Proc. Natl Acad. Sci. USA* **110**, 40–45 (2013).
30. Zhang, J. et al. Integrated spoof plasmonic circuits. *Sci. Bull.* **64**, 843–855 (2019).
31. Gao, X. et al. Programmable multifunctional device based on spoof surface plasmon polaritons. *IEEE Trans. Antennas Propag.* **68**, 3770–3779 (2020).
32. Zhang, H. C., Cui, T. J., Zhang, Q., Fan, Y. & Fu, X. Breaking the challenge of signal integrity using time-domain spoof surface plasmon polaritons. *ACS Photonics* **2**, 1333–1340 (2015).
33. Gao, X. et al. Programmable surface plasmonic neural networks for microwave detection and processing. *Nat. Electron.* **6**, 319–328 (2023).
34. Joy, S. R., Erementchouk, M., Yu, H. & Mazumder, P. Spoof plasmon interconnects—communications beyond RC limit. *IEEE Trans. Commun.* **67**, 599–610 (2019).
35. Koenig, S. et al. Wireless sub-THz communication system with high data rate. *Nat. Photonics* **7**, 977–981 (2013).
36. Yang, Y. et al. Terahertz topological photonics for on-chip communication. *Nat. Photonics* **14**, 446–451 (2020).
37. Tripodi, L. et al. Broadband CMOS millimeter-wave frequency multiplier with Vivaldi antenna in 3-D chip-scale packaging. *IEEE Trans. Microw. Theory Tech.* **60**, 3761–3768 (2012).
38. Park, J-D., Thyagarajan, S. V., Alon, E., Kang, S., & Niknejad, A. M. A 260 GHz fully integrated CMOS transceiver for wireless chip-to-chip communication. in *Proc. 2012 Symposium on VLSI Circuits Digest of Technical Papers*, 48–49 (2012).
39. Zhou, T. et al. In situ optical backpropagation training of diffractive optical neural networks. *Photonics Res.* **8**, https://doi.org/10.1364/prj.389553 (2020).
40. Ako, R. T., Upadhyay, A., Withayachumnankul, W., Bhaskaran, M. & Sriram, S. Dielectrics for terahertz metasurfaces: material selection and fabrication techniques. *Adv. Opt. Mater.* **8**, https://doi.org/10.1002/adom.201900750 (2019).
41. Fan, K. et al. Nonlinear terahertz metamaterials via field-enhanced carrier dynamics in GaAs. *Phys. Rev. Lett.* **110**, 217404 (2013).
42. Le Zhang, Q., Chen, B. J., Shum, K.-M. & Chan, C. H. Ultra-wideband and compact terahertz planar load based on spoof surface plasmon polaritons with nickel. *IEEE Trans. Circuits Syst. II* **68**, 1922–1926 (2021).

Acknowledgements

The work is supported in part by the University Grants Committee/Research Grants Council of the Hong Kong Special Administrative Region, China under Grant AoE/E-101/23-N (C.H.C.); the National Natural Science Foundation of China under Grants 92167202 (Q.M.), 62301147 (T.J.C.), and 62288101(T.J.C.); the Major Project of Natural Science Foundation of Jiangsu Province under Grants BK20212002 (T.J.C) and BK20210209 (T.J.C.); and the Fundamental Research Funds for the Central Universities 2242023K5002(T.J.C.); the Natural Science Foundation of Jiangsu Province under Grant BK20230822 (Q.M.).

Author contributions

Q.M., T.J.C., and C.H.C. initiated the plan and supervised the entire study. X.G., Z.G., and Q.M. conceived the idea of this work. X.G. carried out the structural design, and X.G., Z.G., and W.Y.C carried out the simulations. B.J.C. fabricated the samples. X.G. and K.-M. S. carried out the measurements. X.G., Z.G., Q.M., and J.W.Y carried out the data analyses and developed the enhanced experimental protocol. X.G., Q.M., Z.G., T.J.C., and C.H.C. prepared the manuscript with input from all authors. All authors discussed the research. All authors discussed the research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-51210-2.

Correspondence and requests for materials should be addressed to Qian Ma, Tie Jun Cui or Chi Hou Chan.

Peer review information *Nature Communications* thanks Ai Qun Liu and Hanyu Zheng for their contribution to the peer review of this work. A peer review file is available.