



Multilayer nonlinear diffraction neural networks with programmable and fast ReLU activation function

Received: 17 March 2025

Accepted: 13 October 2025

Published online: 24 November 2025

 Check for updates


Yu Ming Ning^{1,2,3,5}, Qian Ma^{1,2,5} , Qiang Xiao^{1,2}, Xin Xin Gao^{1,2},
Qian Wen Wu^{1,2}, Ze Gu^{1,2}, Rui Si Li^{1,2}, Long Chen^{1,2}, Jian Wei You^{1,2} &
Tie Jun Cui^{1,2,4} 

Optical diffractive neural networks are emerging for improving speed and energy efficiency in machine learning. However, the challenges of nonlinear activation functions (e.g., latency issues, high power consumption, and cascading complexity) impede their performance and practical deployment. Here, we propose a programmable multilayer full-space nonlinear neural network operating in the microwave frequency band. Its nonlinear layers are constructed using programmable metasurfaces integrated with RF components, implementing a ReLU-like activation function. The nonlinear architecture achieves a nanosecond-scale delay (17.7 ns), representing orders of magnitude improvement in speed over photoelectric conversion-based nonlinearities. Moreover, the nonlinearity is characterized by exceedingly low thresholds and reconfigurable nonlinear activation functions. The system demonstrates remarkable classification capability in image classification and real-time human posture recognition tasks. Characterized by low latency, high speed, low power consumption, and flexible nonlinear activation, this architecture holds great promise for applications in security screening, medical rehabilitation, human-computer interaction, and numerous other fields.

The surge in neural-network scale precipitates an exponential rise in energy expenditure and training cost, underscoring the need for more efficient alternatives¹. Optical computing stands out as an effective substitute for electronic computation devices, particularly in the artificial intelligence applications, in which optical neural networks demonstrate marked superiority in speed and energy efficiency over their electronic counterparts^{2–13}. Passive optical diffractive neural networks (DNNs) have faced challenges in programmability, whereas programmable metasurfaces provide real-time programmability^{14–16}, which are crucial for applications like computational imaging^{17,18} and adaptive intelligent devices¹⁹. Recent advancements in DNNs in the microwave band include the programmable diffractive deep neural networks²⁰, dynamic holography²¹, planar diffractive deep neural

networks²², and surface plasmon polariton neural networks^{23,24}. These studies highlight the microwave DNN's inheritance of the key strengths in optical computing, including the high-speed processing, parallelism, and energy efficiency.

Considering that the normal diffraction process is linear, how to achieve efficient and high-speed nonlinear activation is a key in the research of DNNs. Hence, significant research efforts have been directed towards developing DNNs that are integrated with the nonlinear activation functions^{25–32}. However, these nonlinearities face substantial practical limitations. For instance, the nonlinearity derived from semiconductor quantum dots (QDs)³³ is inherently constrained by the conversion efficiency of QDs, which is characterized by low transmission and significant energy attenuation, impeding the

¹State Key Laboratory of Millimeter Waves and Institute of Electromagnetic Space, Southeast University, Nanjing, China. ²Institute of Electromagnetic Space, Southeast University, Nanjing, China. ³ZhangJiang Laboratory, Shanghai, China. ⁴Suzhou Laboratory, Suzhou, China. ⁵These authors contributed equally: Yu Ming Ning, Qian Ma.  e-mail: maqian@seu.edu.cn; tjcui@seu.edu.cn

cascading of multiple nonlinear layers. Optical nonlinear activation mechanisms, such as the photonic saturable absorbers^{29,34} and saturation of radiofrequency amplifiers³⁵, exhibit high thresholds and necessitate high-energy inputs to elicit the nonlinear responses. The conventional CMOS sensor-based photoelectric nonlinearities^{36,37} constrained by exposure duration exhibit millisecond (ms) or microsecond (μ s) level latency, which impede the high-speed computational performance. Furthermore, the nonlinearities that rely on analog-to-digital converters (ADCs), digital-to-analog converters and microcontroller units (MCUs)²³ demand complex peripheral devices, which significantly limit the system's computation rate due to the processing constraints of digital circuits. Thus, these issues further impede the capability of DNNs to demonstrate systematic advantages over traditional digital computations in practical tasks.

To address these challenges, here we propose a multilayer nonlinear diffractive neural network (MN-DNN) with the programmable and fast Rectified Linear Unit (ReLU) activation function operating at 5.8 GHz. The network features multiple nonlinear layers composed of nonlinear metasurfaces integrated with RF detectors, amplifiers, and voltage adders, which exhibit the intensity-dependent nonlinear transmission coefficients, and hence can effectively act as the ReLU activation function. By selecting an optimized high-speed device, the engineered single-layer nonlinear metasurface achieves nanosecond-scale delays (17.7 ns), representing a substantial reduction compared to the millisecond (ms) and microsecond (μ s) latency that is typically associated with conventional nonlinear optoelectronic conversion processes^{26,36,37}. The ReLU thresholds can be adjusted through diverse bias voltages, liberating the nonlinear activation from depending on high-intensity illumination, and enhancing adaptability across varied intensity conditions. The gain of nonlinear metasurfaces is controllable via bias voltage, mitigating energy loss, and stacking of multiple nonlinear layers. We show that MN-DNN can enhance the image classification on Modified National Institute of Standards and Technology (MNIST) and Fashion-MNIST datasets, with the accuracy of 92.81% on MNIST and 78.8% on Fashion-MNIST, surpassing the linear networks by over 4% in both cases. It can also directly classify the human postures from electromagnetic waves in real time with the accuracy of 93.06%, surpassing the linear network's accuracy of 81.85%. The advantages of this network in low latency and high accuracy make it ideal for applications in real-time sensing, motion detection, wireless communication, radar signal processing, and beyond.

Results

Principle of MN-DNN

To address the challenges of nonlinear activation, we propose an MN-DNN featuring low latency and adjustable nonlinearity for fast and accurate computations. The traditional DNNs require digital encoding of the input data, which increases the complexity and latency. Our architecture transcends the conventional encoding and enables the processing of both encoded input data and direct electromagnetic wave information, such as for real-time human posture classification shown in Fig. 1a. The direct processing significantly enhances the computational speed and substantially simplifies the hardware. The MN-DNN architecture comprises linear and nonlinear layers, and the latter employs the metasurfaces that incorporate RF amplifiers, detectors, and voltage adders to emulate the ReLU nonlinear activation function based on the power between input and output. Additionally, the implementation of diverse activation functions such as Hyperbolic Tangent (Tanh) and Leaky Rectified Linear Unit (Leaky-ReLU) is straightforward with the alternative active circuit configurations, as elaborated in the "Discussion" section.

In addition to real-time processing advantages, the MN-DNN architecture also exhibits strengths in its programmability and low latency. The nonlinearity is tunable via external bias voltage, enabling adjustments to the activation threshold and slope, as shown in Fig. 1b.

This flexibility eliminates dependence on high power, broadening its applicability across various scenarios. Low latency is essential for swift information processing, but the nonlinearities in the current neural networks often introduce millisecond to microsecond delays owing to their reliance on optoelectronic conversion^{23,26,36,38}, hindering the potential for computation at the speed of light. However, our architecture achieves a latency of 17.7 ns (Fig. 1c), marking a substantial reduction in the delay and underscoring its importance for real-time processing. Hence, MN-DNN boosts the information extraction capabilities, outperforming the linear networks with over 4.2% higher accuracy in both classifications of MNIST and Fashion-MNIST datasets, as depicted in Fig. 1d, e. The MN-DNN also enables real-time extraction of information from electromagnetic (EM) waves. As shown in Fig. 1f, it accurately classifies eight postures with the 93.06% accuracy, enhancing the computational efficiency and reducing the hardware complexity over the conventional methods. Hence, it can be used in human-computer interaction, intelligent surveillance, and industrial inspection.

Architecture of nonlinear metasurface unit

In MN-DNN, the nonlinear layer is implemented with a nonlinear metasurface. The nonlinear unit structure comprises four metal layers and three dielectric layers. The three dielectric layers, from top to bottom, are as follows: 0.5 mm-thick F4B (relative permittivity $\epsilon_r = 2.65$), 0.2 mm-thick FR4 (relative permittivity $\epsilon_r = 4.4$), and another 0.5 mm-thick F4B (relative permittivity $\epsilon_r = 2.65$). The RF energy is coupled by the receiving antenna, passes through the active circuit, and is radiated by the transmitting patch antenna into the air, as illustrated in Fig. 2a.

The front-side passive structure comprises an octagonal receiving patch antenna and a T-junction unequal power divider, as shown in Fig. 2b. Optimized for operations at 5.8 GHz, the patch antenna exhibits a side length dimension of $D = 7.57$ mm to achieve proper resonance characteristics. For impedance matching, the octagonal patch antenna is slotted with a length $L_1 = 4.91$ mm and a width $W_1 = 1$ mm. The microstrip line width W_2 is 1 mm. The detection arm of the T-junction features a two-stage impedance transformation, refined by the corner chamfering to ensure a smoother impedance transition. The first branch has a width $W_3 = 0.5$ mm with length $L_3 = 1.746$ mm, while the second branch has a width $W_4 = 0.254$ mm with length $L_4 = 1$ mm. The unit's back-side features an octagonal transmitting patch antenna identical in size to the receiving antenna, as illustrated in Fig. 2b. The operational diagram of the nonlinear metasurface unit is shown in Fig. 2c. The spatial RF energy, post-coupling by the receiving antenna and microstrip line transmission, is split by the T-junction power divider. The bulk of energy is amplified by the RF amplifier and proceeds to the transmitting antenna, while a fraction of the energy is diverted to a high-speed RF detector and is converted into direct current (DC) voltage V_{OUT} . The combined voltage of V_{OUT} and external bias voltage V_{BIAS} is summed by a voltage adder: $V_{CTRL} = V_{OUT} + V_{BIAS}$. This control voltage is used to modulate the amplifier. Details of the nonlinear metasurface unit structure are presented in Supplementary Note 1.

The RF detector's output voltage V_{OUT} is positively correlated with the input RF energy, as illustrated in Fig. 2d. The gain of RF amplifier is modulated jointly by DC voltages V_{CC} and V_{CTRL} , as shown in Fig. 2e. The effect of DC voltages V_{CC} on the RF amplifier's gain is presented in Fig. 2f. We note that the gain falls below -15 dB when the voltage is less than 0.5 V, effectively muting the amplifier with negligible energy transfer. Above the threshold, increasing V_{CC} will shift the gain into the amplification zone and reach a stable gain. Fig. 2g indicates that low V_{CTRL} values result in a transmission coefficient below -10 dB, halting the transmission. Above 1.2 V, V_{CTRL} initiates the signal amplification, maintaining a stable gain despite further increases. With a high external bias voltage V_{BIAS} (e.g., $V_{BIAS} = 1.5$ V), the amplifier's V_{CTRL}

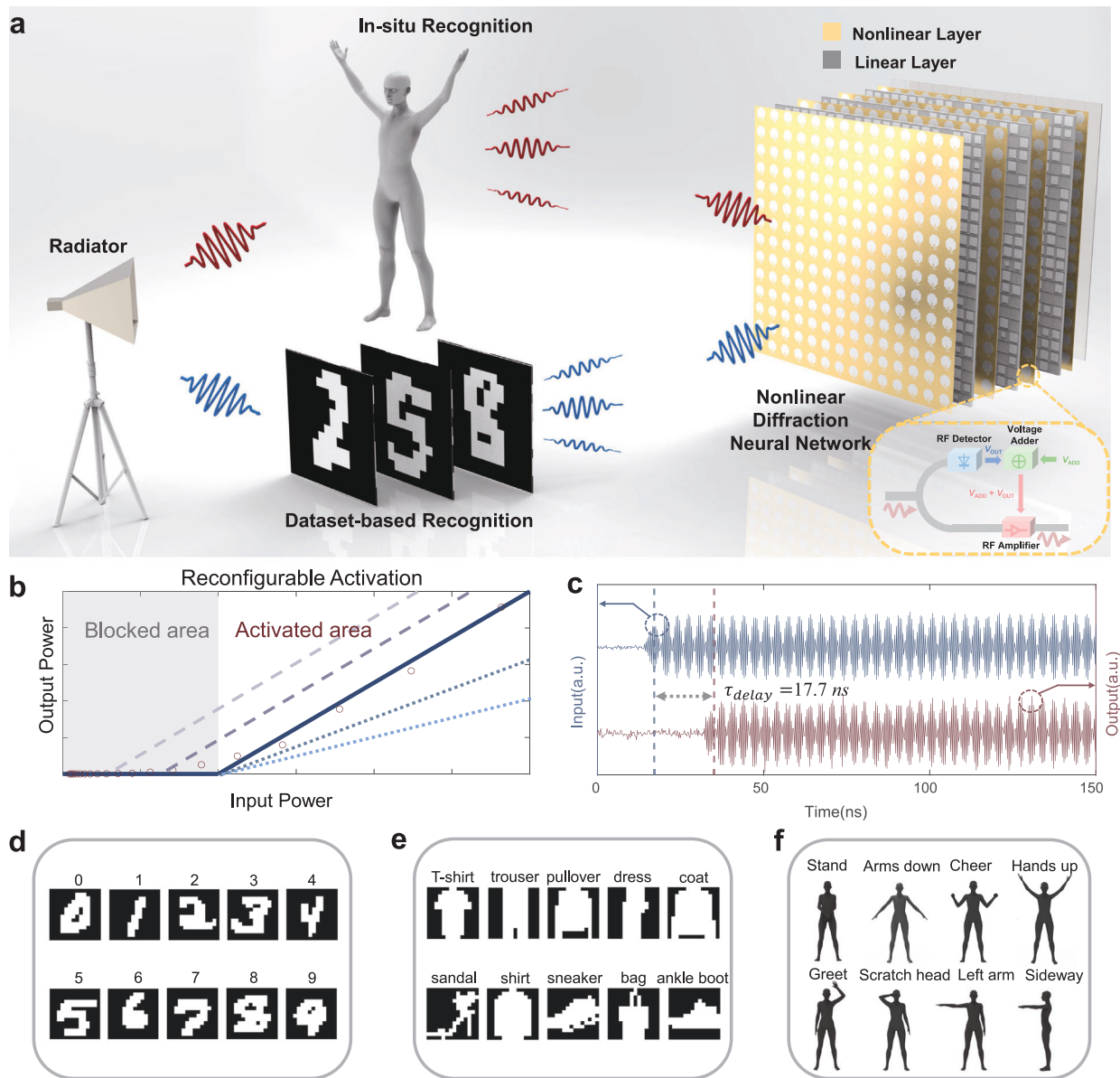


Fig. 1 | Multi-task processing with MN-DNN. **a** Schematic of the MN-DNN operation. The network is able to perform both dataset-based image recognition and real-time human posture classification. Incident EM waves, encoded with input information, are processed by the MN-DNN comprising three linear and three nonlinear layers. Each nonlinear unit integrates an RF detector, an amplifier, and a voltage adder. The classification result corresponds to the location of the

maximum energy on the output plane. **b** Input-output response of the nonlinear metasurface unit, exhibiting a ReLU-like characteristic with a dynamically adjustable threshold and slope. The output is negligible for weak inputs and increases linearly beyond the threshold. **c** Experimentally measured time delay of a single nonlinear metasurface layer (17.7 ns). Experimental demonstration of classification capabilities on the MNIST (**d**), Fashion-MNIST (**e**) and human postures (**f**) tasks.

surpasses the amplification threshold, independent of the detector's output voltage V_{OUT} . As shown in Fig. 2h, In the linear amplification regime, the output power of the metasurface unit is linearly proportional to its input, acting as a linear amplitude modulator with V_{CC} governing the amplitude of the transmission coefficient.

With low V_{BIAS} , the unit is operated nonlinearly, and its transmission coefficient changes with incident power. By increasing the incident power, the detector's V_{OUT} and the amplifier's control voltage V_{CTRL} rise linearly. The amplifier gain rises from below -10 dB to a stable level, and the metasurface unit's transmission coefficient increases from a low value to a stable state. Higher V_{BIAS} leads to an earlier transition into the linear amplification, while lower V_{BIAS} requires more input power to reach the turning point of transmission coefficient, as shown in Fig. 2i. The transmission coefficient S_{21} escalates with V_{CC} magnitude under constant input power and V_{BIAS} , as

illustrated in Fig. 2j. We note that the detector's V_{OUT} and the amplifier's V_{CTRL} are low when the incident power on the metasurface is below 1 mW , leading to amplifier cutoff and zero output from the metasurface unit. Beyond 1 mW , the voltages will be increased, activating the amplifier's linear amplification and causing the metasurface unit's output power to rise linearly with the input power, which emulates the ReLU function in neural networks, as demonstrated in Fig. 2k. In this scenario, V_{BIAS} is 0.5 V and V_{CC} is 5 V . Figure 2l indicates that higher V_{BIAS} reduces the nonlinear activation function's threshold, enabling variable nonlinear activation thresholds by modulating the bias voltage V_{BIAS} . Figure 2m shows that increasing V_{CC} steepens the slope of the nonlinear activation function, resulting in elevated amplifier gain and transmission coefficients for the metasurface units. Based on the measured data (details of the nonlinear metasurface tests are provided in Supplementary Note 2), we construct scatter plots and

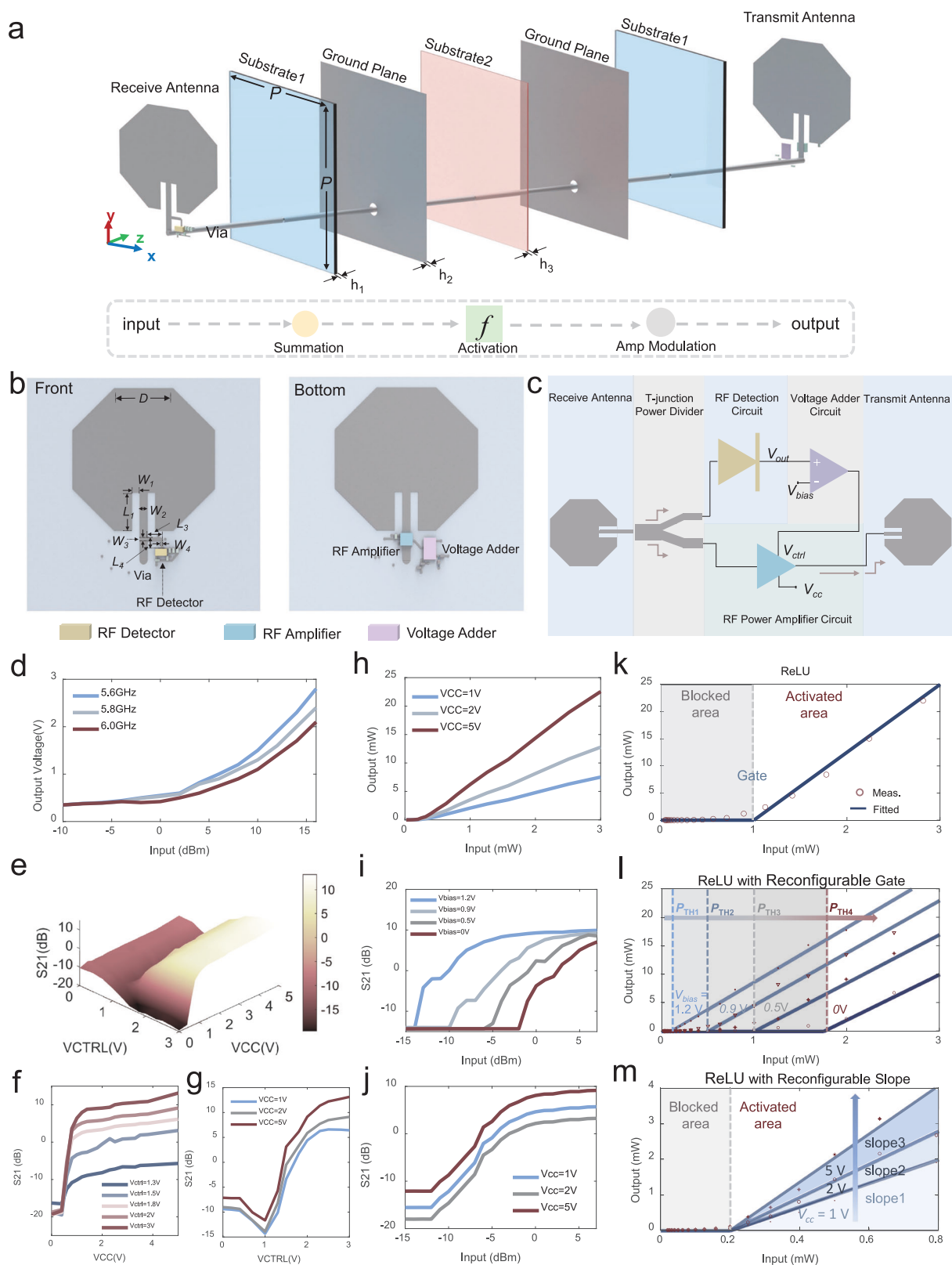


Fig. 2 | Architecture and operational characteristics of the nonlinear metasurface unit. **a** Unit structure and functional equivalence. **b** Front and back views of the unit. **c** Operational schematic. **d** RF detector's output DC voltage increases with incident power density. **e** RF amplifier gain is controlled by DC voltages V_{CC} and V_{CTRL} . Gain response versus V_{CC} (**f**) and V_{CTRL} (**g**), both showing a cutoff region at low voltages, followed by an increase and eventual saturation.

h Linearity of the output energy versus input under high V_{BIAS} ($V_{BIAS} > 1.2$ V), and gain increases with V_{CC} . **i, j** Transmission coefficient rises from cutoff to saturation with incident energy; V_{BIAS} lowers energy threshold (**i**), and V_{CC} raises transmission saturation (**j**). **k** With $V_{BIAS} = 1$ V and $V_{CC} = 5$ V, transmission is zero below 1 mW and increases linearly above it, resembling a ReLU response. **l** Activation threshold decreases with V_{BIAS} . **m** Gain increases with V_{CC} .

employ curve fitting to derive the mathematical expression of the activation function, which exhibits a ReLU-like form:

$$y = b \times \max(0, x - a) \quad (1)$$

where x represents the input power, y denotes the output power, a is the threshold parameter, and b is the slope parameter. This nonlinear activation function offers 17.7 ns latency in the measurements (see Supplementary Note 3 for details). Each nonlinear unit exhibits a total power dissipation of approximately 429 mW, which consists of 150 mW for the RF amplifier, 53 mW for the RF detector, and 224 mW for the voltage adder. Supplementary Note 4 and Supplementary Table S2 compare the performance of diverse nonlinear implementation methods in DNNs, focusing on metrics such as time delay, threshold, multi-layer stacking, and programmability.

Handwritten digits classification by MN-DNN

The primary objective is the classification of MNIST dataset, which involves identifying ten classes of handwritten digits ranging from 0 to 9, as shown in Fig. 3a. The original 28×28 pixel images are down sampled to 14×14 pixels to match the dimension of the input-layer metasurface via bilinear interpolation, and subsequently binarized to values of 0 or 1. Further details on image processing for recognition are outlined in Supplementary Note 5, with the effect of image pre-processing on recognition accuracy analysis provided in Supplementary Note 6. Intensity modulation is achieved by mapping the input image pixels onto the first-layer metasurface. The background is opaque to the EM waves, with the digits being transparent (transmission coefficients of 0 dB for pixel 0 and 10 dB for pixel 1), as illustrated in Fig. 3b. Consequently, the EM wave conveys the information of the input image as it passes through the first layer.

As shown in Fig. 3d, we independently train linear and nonlinear neural networks using the full MNIST dataset (60,000 for training and 10,000 for testing) to evaluate their image classification performance. The linear network comprises three linear phase-modulation layers spaced 20 cm apart, each composed of a 22×22 array of 19×19 mm² units. The phase of each unit serves as a trainable parameter, adjustable in the range from 0 to 2π (see details in Supplementary Note 7). The nonlinear network features an alternating linear-nonlinear structure with three linear layers interspersed with three nonlinear metasurface layers, spaced at 10 cm intervals. Each nonlinear layer contains two independent trainable parameters: the threshold and slope of the ReLU activation function, with layer-specific values (see Supplementary Note 8 for the RF component math modeling). The nonlinear network consists of 1458 trainable parameters, with 484 parameters per linear layer and 2 parameters (threshold and slope) per nonlinear layer, yielding a total of $484 \times 3 + 2 \times 3 = 1458$, and all parameters are trained together. In comparison, the standard fully connected networks for the MNIST classification typically require tens to hundreds of thousands of parameters, whereas our model achieves an order-of-magnitude reduction with only ~1500 parameters. The linear and nonlinear networks remain fully independent throughout their design and training phases, without shared parameters or cross-influence. The loss function is a weighted sum of the Mean Squared Error and SoftMax Cross-Entropy, with coefficients 0.4 and 0.6, respectively. The gradient descent optimization is performed using the Adam algorithm. The mathematical model of DNNs is detailed in Supplementary Note 9. The optimization procedure explicitly accounts for the field distribution of 5.8 GHz horn antenna on the input plane under the test conditions (Supplementary Note 10), to reduce the discrepancies between the simulations and measurements. The training outcomes indicate that the nonlinear network achieves a recognition accuracy of 92.81%, which is 4.18% higher than the 88.63% accuracy of the linear network. It is important to note that this improvement does not result from the

increase of trainable parameters, as each nonlinear layer only adds two trainable parameters (threshold and slope), and three nonlinear layers contribute a total of six additional parameters. For comparison, a linear network with the same number of parameters achieves at most 89% accuracy, further confirming the performance gain from the nonlinear activation functions (see Supplementary Note 11). The output fields of both networks are reduced to two dimensions via t-SNE, as shown in Fig. 3e. The clustering of the linear network shows a looser intra-cluster structure with overlapping data points across categories, whereas the output of the nonlinear network displays higher local density within clusters and clearer separation between categories.

Based on the simulation results, nonlinear and linear network samples are fabricated, in which the metasurfaces are mounted on acrylic plates with plastic screws and aligned into the grooves of aluminum alloy brackets. Absorptive materials are applied to the plates and brackets to reduce the edge diffraction of EM waves and to minimize their influence on the EM fields, see Supplementary Note 12 for details. The nonlinear prototype includes an input layer, three linear layers and three nonlinear layers of metasurfaces. The input-layer metasurface functions as a configurable linear amplitude modulator in a 0–10 dB range, which is implemented by a nonlinear metasurface in linear amplification mode, controlled by a bias voltage. It consists of 14×14 units, where the transmission coefficient of each unit matches the input image pixel values, modulated by an MCU-regulated bias voltage. The activation function parameters of nonlinear metasurfaces are modulated by the external DC voltage. The linear prototype comprises an input-layer metasurface and three linear phase-modulating metasurfaces, configured analogously to the nonlinear network. Both linear and nonlinear prototypes are excited by a 5.8 GHz horn antenna.

The input-layer encoding was performed through intensity modulation. As depicted in Fig. 3c, a two-dimensional near-field scanning platform recorded the output field distributions. To comprehensively evaluate the network performance, we systematically assessed the recognition accuracy of both networks using 5000 randomly selected images. In the experimental setup, a rapid detection device was used to replace the conventional two-dimensional near-field scanning stage (see Supplementary Note 13), enabling efficient energy measurements across ten specific regions of the output field for each image. This approach significantly enhanced the detection throughput, allowing rapid acquisition of classification results for large-scale image datasets. The MNIST rapid classification experiments conducted with this device are presented in Supplementary Video 1. The experimental results demonstrate that the nonlinear network achieves a classification accuracy of 92.6% on the 5000 MNIST images (compared to a simulated accuracy of 92.81%), exhibiting a clear advantage over the linear network (88.5% accuracy), as shown in Fig. 3f. To further analyze the networks' behavior, we performed output field distribution scans on a subset of images. For simple images with distinct features such as the handwritten digit "7" (Fig. 3g), both linear and nonlinear networks achieve accurate recognition, with their output energies precisely concentrated on the target category's region. However, when processing complex images with easily confusable features, the nonlinear network shows a pronounced advantage. As seen in Fig. 3h, for a "4" sample prone to confusion with "9", the nonlinear network's output energy is correctly localized to the '4' region, whereas the linear network misclassifies it, with its peak energy erroneously appearing in the "9" region, although the "4" region has the second-highest energy. This phenomenon remains consistent across other complex image classifications, demonstrating the nonlinear network's superior performance in handling complex images. Further comparisons are presented in Supplementary Note 14. A detailed comparison between simulated and experimental results is given in Supplementary Note 15. The total power of MN-DNN is about 252.352 W (see Supplementary Note 16 for details).

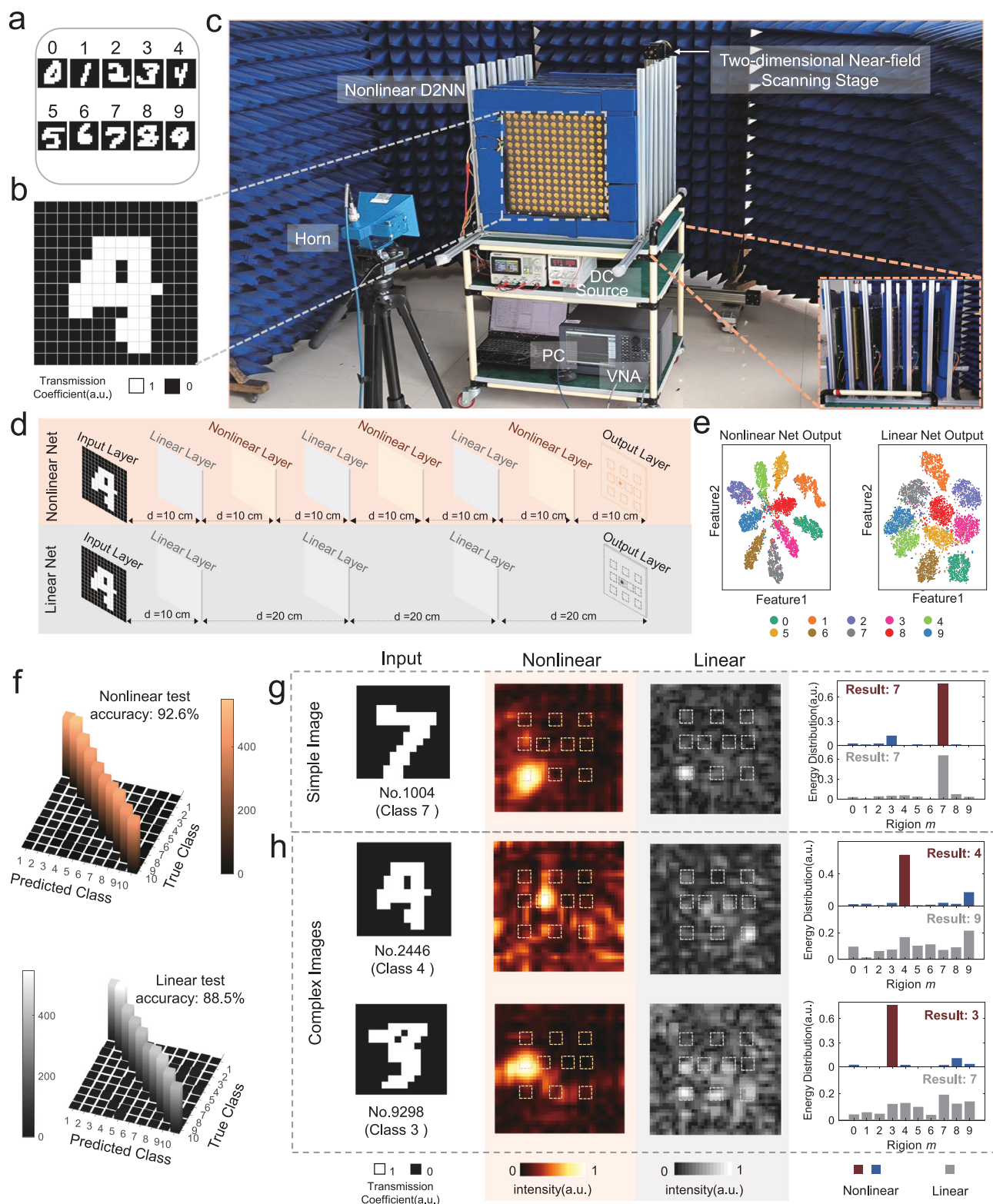


Fig. 3 | Handwritten digits classification. **a** Example images from the MNIST dataset, which comprises 10 digit classes (0–9). **b** Encoding binarized images into the transmission coefficients of the input metasurface. **c** Experimental test setup. **d** Schematic of the two network architectures: a nonlinear network with interleaved linear and nonlinear layers, and a linear network solely with linear layers. **e** t-SNE visualization of test set outputs for both networks, showing clearer class separation and tighter clustering in the nonlinear network. **f** Confusion matrices show 92.6%

(nonlinear) and 88.5% (linear) accuracy on 5000 test images. **g** Sampled output field distributions generated by the linear and nonlinear networks for a simple image, demonstrating that both networks focus on the target region. **h** Sampled output field distributions for two complex images from both networks, in which the nonlinear network focuses on the correct regions while the linear network focuses on incorrect regions.

Fashion-MNIST classification by MN-DNN

Next, we compare the classification performance of linear and nonlinear networks on the complete Fashion-MNIST dataset, which consists of 60,000 training samples and 10,000 test samples. This dataset comprises ten apparel categories: T-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots, as illustrated in Fig. 4a. The 28×28 pixel images are resized to 14×14 pixels to match the input layer metasurface using bilinear interpolation and then binarized. Intensity modulation is applied by mapping the pixel values onto the first-layer metasurface. To assess the image classification capabilities of linear and nonlinear neural networks, we optimize both structures to map the output field distributions of ten classes to ten distinct regions, as shown in Fig. 4b. The network architectures for Fashion-MNIST recognition mirror to the architectures employed in MNIST, and both linear and nonlinear networks undergo independent optimization processes without any parameter sharing. The nonlinear diffraction network achieves 78.8% accuracy on Fashion-MNIST, surpassing the linear network (74.43%) by 4.37 percentage points. Similar to the previous case, this improvement is not due to the increased parameter count, as each nonlinear layer only introduces two trainable parameters (the threshold and slope) with a total of six additional parameters, which are negligible compared to the model's 1458 parameters. Under the same parameter conditions, the linear network achieves 75.08% accuracy (Supplementary Note 11), further demonstrating the performance gain from the nonlinear architecture. The accuracy comparison with other networks is given in Supplementary Note 17. Based on the simulations, we fabricate the nonlinear and linear prototypes. During experimental evaluation, a set of 5000 randomly sampled Fashion-MNIST test images is intensity-encoded and processed through both nonlinear and linear network configurations. The recorded recognition accuracy reaches 77.4% and 72.5% for the nonlinear and linear networks, respectively. The resulting output field distributions of linear and nonlinear samples from the horn antenna irradiation are shown in Fig. 4c, d. For simple and distinct images like the ankle boot in Fig. 4c, both linear and nonlinear networks achieve accurate classification, with the output energy correctly focused on the target region. However, when dealing with complex images characterized by similar and confusable features, the nonlinear network demonstrates a clear advantage. As illustrated in Fig. 4d, the nonlinear network correctly classifies the sneaker image (often confused with ankle boots), with the output energy concentrated specifically in the sneaker region. In contrast, the linear network misclassifies this input, displaying the peak energy in the wrong ankle boot region despite the secondary activation in the correct sneaker category. This consistent pattern across challenging images (see Supplementary Note 18 for more results) confirms the nonlinear networks' superior performance in recognizing the complex images.

In addition to the experiments, we conduct a series of simulation analyses to determine the impact of the number, placement, and type of activation functions in nonlinear layers on the recognition accuracy. We assess different network configurations with varying numbers of nonlinear layers: a three-layer purely linear network; a network with an additional nonlinear layer between the first and the second linear layers; a network with two nonlinear layers interspersed among the linear layers; and a network with three nonlinear layers evenly distributed among the linear layers. The results show that incorporating nonlinear layers can significantly enhance the recognition accuracy over the purely linear network, with the accuracy increasing as the number of nonlinear layers grows, as illustrated in Fig. 4e. In the three-linear-layer framework, we introduce nonlinear layers at three different locations: between the first and the second linear layers (front), between the second and the third linear layers (middle), and between the third linear layer and the output plane (rear). The results illustrate that the recognition accuracy increases with the addition of nonlinear layers at all positions, with the largest improvements at the front and

rear, as depicted in Fig. 4f. In a three-linear-layer structure, we introduce a nonlinear layer preceding the output plane to test the ReLU, SoftMax, and Tanh activation functions. Each function boosts the accuracy significantly, with ReLU showing the greatest enhancement, as depicted in Fig. 4g. For more details, refer to Supplementary Note 19.

Static posture recognition by MN-DNN

We develop a posture recognition system using MN-DNN, as shown in Fig. 5a. A test person stands in front of MN-DNN and performs various postures, which affect the EM-wave focusing on the output plane. Our classification targets include eight distinct human body postures: stand, arms down, cheer, hands up, greet, scratch head, left arm, and sideway. Firstly, we collect a dataset of the postures. The system employs a standard horn antenna operating at 5.8 GHz to illuminate the test person. Posture information is extracted using an eight-element patch antenna array with 3-2-3 arrangement (5.8 GHz resonant frequency) that samples the scattered field distributions from the human body. Signals from these antennas are sequentially captured by Vector Network Analyzer (VNA), managed by a computer that oversees an MCU to control a single-pole eight-throw RF switch (refer to Supplementary Note 20). Four volunteers are designated as trainers and perform extensive data acquisition for eight postures under various conditions, as depicted in Supplementary Fig. S42. These include different rotation angles of human body relative to the antenna array (ranging from -30° to 30° in 10° increments) and varied distances between the human body and antenna array (from 20 cm to 50 cm in 10 cm increments), as presented in Supplementary Note 21. The input layer of our system features 8 active units with a transmission coefficient of 10 dB, arranged in a 3-2-3 array configuration, aligning with the antenna array's positioning during the dataset acquisition. These units are tasked with spatially sampling the scattered EM waves emitted by the human body. The remaining input layer units are inactive, blocking the transmission of corresponding EM waves, as illustrated in Fig. 5a. The strategic selection of 8 sampling points in the input layer is crucial to attain high recognition accuracy in the limited dataset acquisition timeframe (see Supplementary Note 22 for more details). A total of 850 data samples are collected for each posture, resulting in 6800 samples across eight postures, with corresponding labels ranging from 0 to 7. Among these, 70.5% are allocated to the training set, 23.5% to the test set, and 6% to the validation set (see Supplementary Table S9 for more details).

We further demonstrate a low-latency and high-fidelity posture recognition task based on the proposed MN-DNN architecture. In traditional methods, the input data streams are processed in the digital domain, which has limitation in processing speed and requires complex front-end analog-to-digital conversion modules. In the proposed MN-DNN, the scattered field undergoes neural network computation during its propagation through the network, thereby overcoming these limitations. The MN-DNN comprises three linear layers and three nonlinear layers, as previously described and depicted in Fig. 5b. The output field distributions for various postures are mapped to eight distinct regions on the output plane. Simulation results show that the nonlinear network can quickly achieve a recognition accuracy of 93.06%, which substantially surpasses the linear network's 81.85%, as detailed in Supplementary Note 23.

A trainer is invited to perform the test, as depicted in Fig. 5c. The EM intensity across eight specific regions on the output plane corresponds to the classification of eight distinct postures. In each region, a patch antenna coupled with an RF detector is utilized to rapidly measure the energy of the area, converting it linearly into DC voltages (V_1 – V_8). The outputs from these detectors are connected to the ADC ports of the MCU for voltage detection. MCU compares the voltage magnitudes and controls the color of LEDs based on the channel with the maximum voltage. Each posture classification is associated with a unique LED color: “stand”—white, “arms down”—red, “cheer”—yellow,

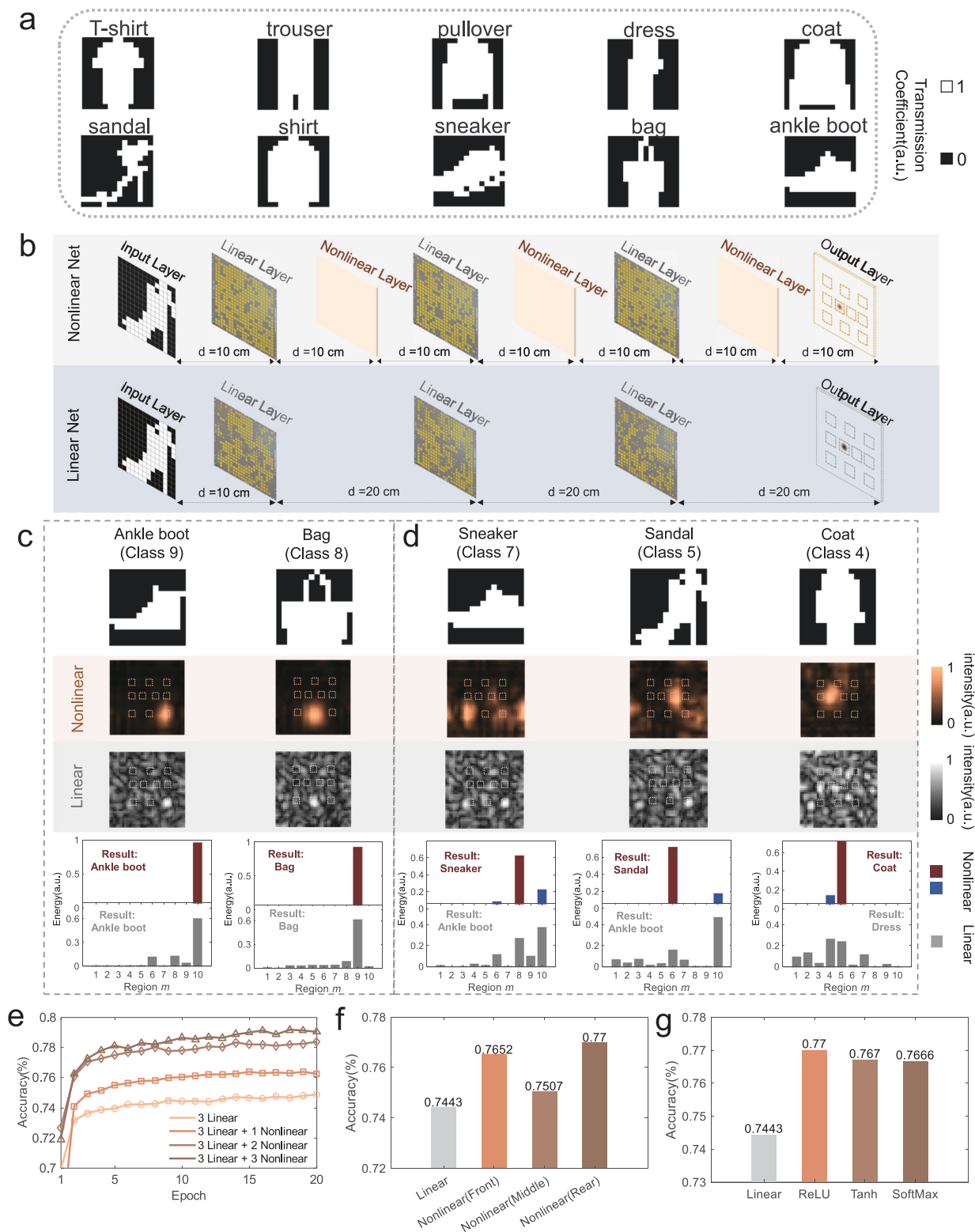


Fig. 4 | Fashion-MNIST classification. **a** The 10-class fashion item dataset for recognition. **b** Comparison of two network architectures: the nonlinear network with alternating linear and nonlinear layers, and the linear network with only linear layers. **c** Sampled output field distributions generated by the linear and nonlinear networks for two simple images. Both networks successfully focus on the target regions. **d** Sampled output-field distributions for three complex images from both

networks. The nonlinear network focuses on correct regions and the linear network on incorrect regions. **e** Recognition accuracy increases with the number of nonlinear layers. **f** Effect of nonlinear layer positioning on the accuracy, with the front and back placements outperforming the central insertions. **g** Impact of activation function type. The ReLU function provides the most significant performance improvement.

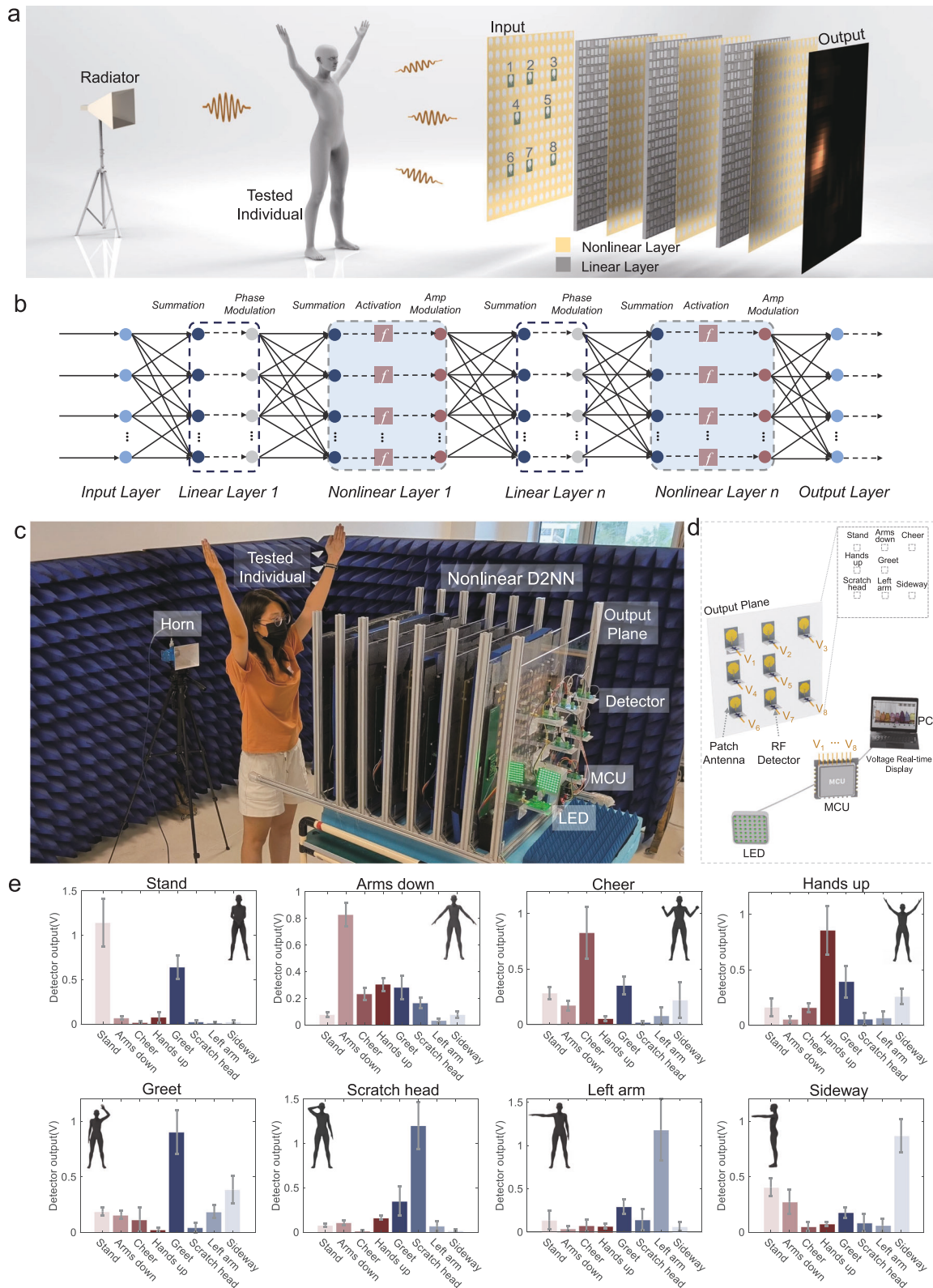


Fig. 5 | Posture recognition. **a** Schematic of MN-DNN for posture recognition, where the scattered EM fields from the test person are processed by MN-DNN. **b** Equivalent diagram of MN-DNN. **c** Experimental setup for posture recognition. **d** Schematic of the detection setup on the output plane, consisting of eight zones each equipped with an antenna and detector that convert EM energy into DC

voltages (V_1 – V_8). These voltages are processed by ADC of MCU, controlling the LED colors based on the maximum voltage channel, and transmitted to a computer for real-time display. **e** Experimental results from 50 repetitions of eight postures, showing the accurate classification as confirmed by voltage readings across the eight regions.

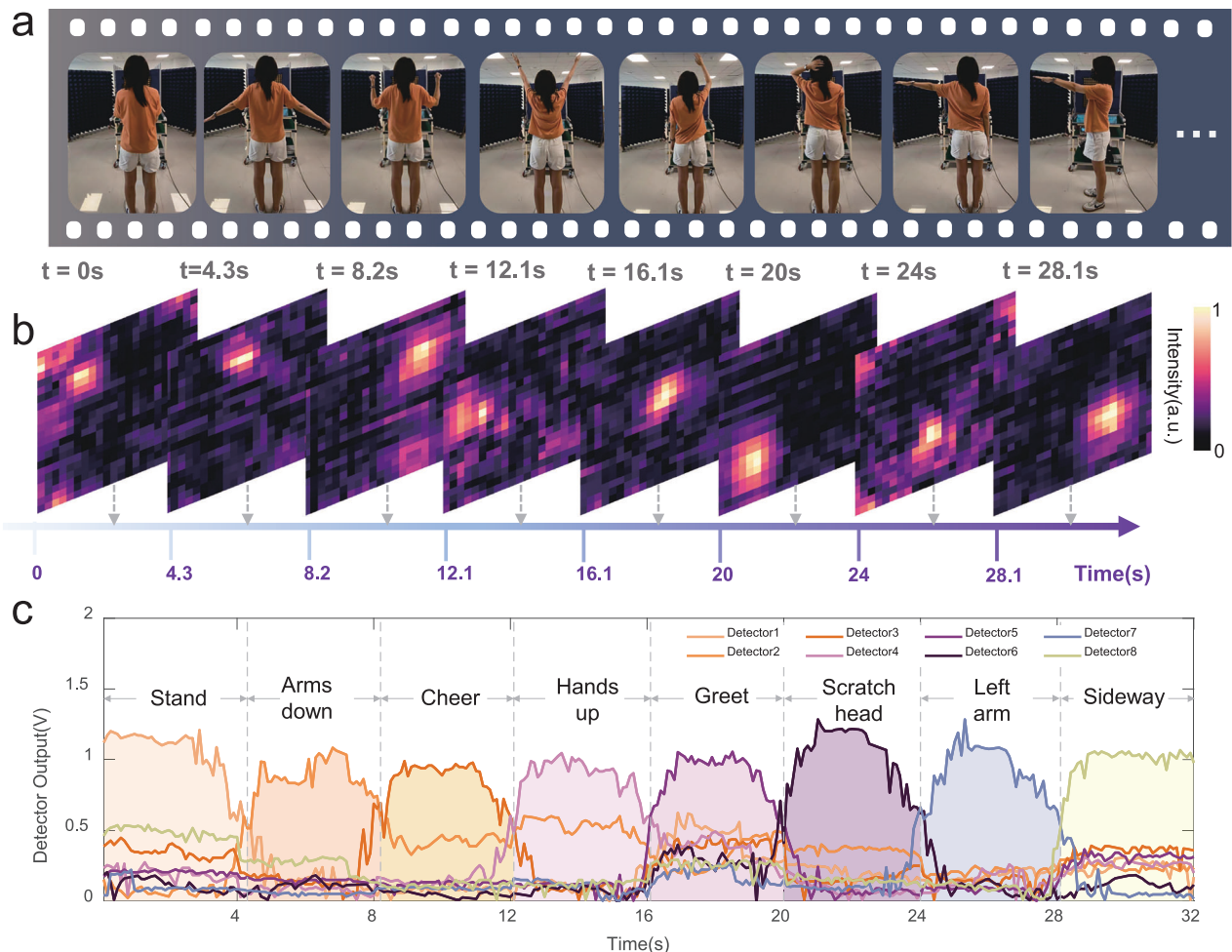


Fig. 6 | Dynamic posture recognition with MN-DNN. a A tester performs eight distinct postures in front of the MN-DNN, switching approximately every 4 s. **b** Simulated output field distributions for the corresponding postures, with the focal points progressively moving across eight designated regions. **c** Real-time

detected voltage waveforms from the eight regions on the output plane. During each posture, the voltage in the corresponding region is higher than that in the others, indicating the correct classification.

“hands up”—green, “greet”—blue, “scratch head”—light blue, “left arm”—purple, and “sideways”—orange. Additionally, MCU conveys the voltage values of the eight channels to a personal computer for real-time display (see Fig. 5d). The underlying principles of energy detection apparatus on the output plane are elaborated in Supplementary Note 24.

The tester performs each of the 8 postures 50 times, and the output voltages from eight RF detectors are logged to measure the intensity across eight distinct regions. Classification is determined by the posture corresponding to the detector with the highest voltage, as shown in Fig. 5e, where the mean output voltage of the detectors is represented by bar heights and their variances by error bar lengths. Consistent with expectations, the regions with the maximum energy on the output plane align with the executed postures. For example, during the “scratch head” posture, Detector 6 records an output voltage nearing 1.2 V, while others remain below 0.5 V, indicating the highest voltage and thus classifying the posture as “scratch head” in accordance with Fig. 5d. More experimental pictures are presented in Supplementary Note 25. To evaluate the system’s robustness, we conduct experiments with four additional participants to exhibit diverse body shapes and clothing in complex backgrounds, none of whom had participated in the initial data collection. The results show an average recognition accuracy of 87%. Future improvements could be focused on expanding the data collection scope by incorporating more sampling points and recruiting additional participants with

varied body shapes, clothing styles, locations, and environmental conditions, which would enhance the system’s adaptability and robustness across diverse scenarios (see Supplementary Note 26 for detailed analysis).

Dynamic posture recognition by MN-DNN

MN-DNN can effectively recognize both static and dynamic postures. The tester performs a sequence of postures, holding each for approximately 4 s before transitioning to the next. The sequence encompasses stand, arms down, cheer, hands up, greet, scratch head, left arm, and sideway, as shown in Fig. 6a. The simulated field distributions corresponding to these postures on the output plane are presented in Fig. 6b, with the field’s focus sequentially through eight regions in chronological order. The output voltages from detectors are interfaced with MCU, which are then transmitted to a computer in real time via serial communication. Figure 6c presents the temporal voltage waveforms. During the interval from 0 to 4.3 s, the posture is “stand”, with Detector 1 exhibiting an output voltage of about 1.2 V, while other detectors show less than 0.55 V, resulting in its classification as “stand”. From 4.3 to 8.2 s, as the posture shifts to arms down, the voltage in Detector 1 drops significantly, and the voltage in Detector 2 increases to approximately 0.9 V, peaking and classifying the posture as “arms down”. This pattern holds for the subsequent postures, ending at the 32-s mark. Throughout each posture’s duration, the relevant detector’s output voltage dominates, allowing for

accurate classification of dynamic movements despite minor outputs from other detectors. The complete tests are given in Supplementary Video 2. The total system latency is 2.48 μ s, comprising both propagation delay through MN-DNN and processing delay incurred by the output detection. This represents an at least three-order-of-magnitude reduction in latency compared to the conventional camera systems coupled with digital neural network post-processing, which are typically operated with millisecond-level latency. For posture recognition tasks, real-time performance typically requires a time delay on the order of milliseconds^{39–41}. The proposed system reduces the latency by at least three orders of magnitude, enabling efficient real-time motion recognition. By implementing a higher-speed output detection scheme with high-speed ADC modules and high-performance FPGA, the total system latency could be further reduced to 66.5–68.5 ns. See Supplementary Note 27 for details.

Discussion

We presented a novel MN-DNN by integrating RF amplifier, RF detector, and voltage adder into the metasurface unit, enabling ReLU-like nonlinearity with nanosecond delay, surpassing the conventional optoelectronic methods. The system latency can be further reduced through multiple optimization approaches (see Supplementary Note 28 for details). By employing high-performance active components in the nonlinear layer, the time delay can be optimized from the nanosecond down to picosecond range. For example, using operational amplifiers with high slew rates can reduce the computational delay to below 200 ps, while using fast-response diodes (e.g. fast Schottky barrier diodes) as RF detectors can shorten the delay to 220 ps. Further improvements can be achieved by replacing the conventional RF amplifiers with high-speed RF Schottky diode switches (\sim 6 ps), enabling superior response speeds. The theoretical calculations indicate that a single-layer delay can be reduced to within 426 ps, which is comparable to the fastest reported nonlinear layers. With a moderate trade-off in reconfigurability, this delay can be further reduced to 226 ps. In addition, the propagation delay can be effectively minimized by optimizing the nonlinear network architecture (e.g., using increased operational frequency and reduced interlayer spacing), thereby reducing the radial distance between the input and output planes. For the type of activation functions, alternative activation functions can be implemented by modifying active circuits of the nonlinear metasurface, including Tanh and Leaky-ReLU, as presented in Supplementary Note 29.

Furthermore, MN-DNN has the potential for in-situ training, as it can adjust the bias voltages of nonlinear layer units via FPGA and extract the output voltages from the integrated detectors to represent node intensity. Future implementations will detect the output plane energy distributions with an RF detector array, followed by gradient computation and weight updates using gradient descent through FPGA for cyclic training. The system can be extended to millimeter-wave and terahertz frequencies by integrating the high-frequency semiconductor devices (e.g., InP/SiGe amplifiers^{42–45} and Schottky diode detectors^{46,47}) with the metasurface designs. This configuration maintains integration density while supporting higher-frequency operation, resulting in highly integrated systems suitable for miniaturization. Improvements in energy efficiency can be achieved by adopting faster and lower-power devices while scaling up the network. Owing to its strong scalability, our network achieves higher power efficiency when scaled to larger single-layer configurations. For instance, the system with a 56×56 -unit single nonlinear layer reaches 4 TOPS/W, positioning it among the state-of-the-art in energy-efficient designs. For image classification tasks, future architectural improvements could target two key aspects to enhance the recognition accuracy and adapt to a broader range of tasks. Firstly, the integration of a digital-to-analog converter in the input-layer

metasurface can enable continuous grayscale inputs, thereby improving information fidelity and subsequent feature extraction accuracy (see Supplementary Note 30). Secondly, systematic scaling of the network dimensions by increasing the metasurface unit density and layers can strengthen the nonlinear processing capacity, ultimately boosting the pattern discrimination performance. In this work, we demonstrated that MN-DNN outperforms the linear networks significantly in image classification accuracy using the MNIST and Fashion-MNIST datasets. We also showed the capability of MN-DNN in real-time processing of spatial EM wave and accurately identifying static and dynamic postures. Owing to its low latency, good adaptability and wide applicability, the proposed MN-DNN holds significant promise for real-time perception, motion recognition, and information processing, indicating a bright future for technological innovation.

Methods

Time delay measurement of a nonlinear metasurface

The time-delay measurement system employs a 5.8 GHz signal generator to produce a continuous radio-frequency (RF) signal. This signal undergoes amplitude-shift keying modulation through a square wave (50% duty cycle) received from a waveform generator via an external trigger interface, thereby generating a modulated RF signal with a square-wave envelope. The modulated signal is then radiated by a transmitting antenna to excite the metasurface. To minimize the edge diffraction effects, we implement a coding strategy that activates only the central unit while deactivating the surrounding edge units, which are also covered with microwave-absorbing material. For real-time monitoring of the incident signal on the metasurface, a receiving antenna (Antenna 1) is positioned equidistant from the transmitting antenna and the central unit of metasurface. The RF signal captured by Antenna 1 is transmitted via a coaxial cable to Channel 1 of an oscilloscope. Simultaneously, a second receiving antenna (Antenna 2), placed on the opposite side of the metasurface, detects the transmitted wave and relays the signal through a coaxial cable of identical length to Channel 2 of the oscilloscope. By analyzing the temporal difference between the signal envelopes of two channels, the transmission delay introduced by the single-layer nonlinear metasurface is precisely quantified. For more details, see Supplementary Note 3.

Data collection for posture recognition

Trainers 1 to 4 conduct comprehensive data collection across eight human postures: stand, arms down, cheer, hands up, greet, scratch head, left arm, and sideway. Positioning themselves between a horn and an antenna array, they perform these postures facing the antenna array. The array captures scattered EM waves, which are then transmitted to a VNA for computer recording. To enhance dataset robustness, data collection encompasses a spectrum of scenarios, including body rotations (from -30° to 30° in 10° increments) relative to the antenna array's normal direction and distances (from 20 cm to 50 cm in 10 cm increments) to the array. A total of 850 datasets are collected for each posture, resulting in an aggregate of 6800 across all postures, which are labeled from 0 to 7. These datasets are randomized and allocated into training (70.5%), testing (23.5%), and validation (6%) sets. Additionally, to further improve dataset robustness, we increase sampling points from 8 to 21. Eight participants with diverse body types and clothing materials perform eight postures repeatedly under both controlled and interference-rich conditions (5.8 GHz WiFi with multipath reflections). As shown in Supplementary Note 26, this enhanced dataset demonstrates significantly improved robustness in simulations.

Analysis of the count and position of nonlinear layers

(1) Number of nonlinear layers. We analyze the influence of nonlinear layer count on recognition accuracy using the MNIST and Fashion-MNIST datasets. The primary focus is on MNIST, where four distinct

neural network configurations were simulated: a purely linear network with three layers, and three nonlinear networks with additional one, two, and three nonlinear layers, respectively. The architectures and simulation outcomes are detailed in Supplementary Fig. S35a–d, with the scatter plots illustrating the output fields post t-SNE dimensionality reduction. These plots reveal that increased nonlinearity enhances output field clustering, with intra-class samples clustering more tightly and inter-class boundaries becoming more distinct. Supplementary Fig. S35e corroborates that recognition accuracy escalates with an increment in nonlinear layers. Supplementary Fig. S36 delineates the Fashion-MNIST recognition outcomes across the four network categories, with accuracy escalating from 74.43 to 78.8%, underscoring a positive correlation between nonlinear layer count and recognition precision. Collectively, the findings from both datasets suggest that augmenting nonlinear layers positively influences image recognition accuracy.

(2) Position of nonlinear layers. We investigate the effect of nonlinear layer positioning on the recognition accuracy using MNIST and Fashion-MNIST datasets. The analysis focuses on four neural networks, each with three linear layers and a single nonlinear layer inserted at different points: inter-first-second, inter-second-third, and post-third linear layers. The detailed architectures and corresponding simulation results on the MNIST dataset are provided in Supplementary Fig. S37a–d. Results demonstrate that the introduction of nonlinear layers at any position improves recognition accuracy, with the highest gains observed when the nonlinear layer is near the output layer. This indicates that the performance is sensitive to the placement of the nonlinear components. Supplementary Fig. S38 shows the classification performance of these network configurations on the Fashion-MNIST dataset, confirming the MNIST dataset trend with the most significant accuracy improvements when the nonlinear layers are positioned near the output layer and the least improvement when placed in the middle.

Data availability

The data supporting the findings of this study are presented in the paper and in the Supplementary information.

Code availability

The code that supports the findings of this study are available from the corresponding author upon request.

References

- Silver, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016).
- Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- Chen, Y. et al. All-analog photoelectronic chip for high-speed vision tasks. *Nature* **623**, 48–57 (2023).
- Xue, Z. et al. Fully forward mode training for optical neural networks. *Nature* **632**, 280–286 (2024).
- Xu, Z. et al. Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence. *Science* **384**, 202–209 (2024).
- Zhu, H. H. et al. Space-efficient optical computing with an integrated chip diffractive neural network. *Nat. Commun.* **13**, 1044 (2022).
- Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).
- Bogaerts, W. et al. Programmable photonic circuits. *Nature* **586**, 207–216 (2020).
- Wang, Z., Chang L., Wang F., Li T., Gu T. Integrated photonic metasystem for image classifications at telecommunication wavelength. *Nat. Commun.* **13**, 2131 (2022).
- Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
- Zheng, Z. et al. Dual adaptive training of photonic neural networks. *Nat. Mach. Intell.* **5**, 1119–1129 (2023).
- Pai, S. et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* **380**, 398–404 (2023).
- Mennel, L. et al. Ultrafast machine vision with 2D material neural network image sensors. *Nature* **579**, 62–66 (2020).
- Cui T. J., Qi M. Q., Wan X., Zhao J., Cheng Q. Coding metamaterials, digital metamaterials and programmable metamaterials. *Light Sci. Appl.* **3**, e218 (2014).
- Ma Q., Cui T. J. Information Metamaterials: bridging the physical world and digital world. *PhotonIX* **1**, (2020).
- Xiao Q. et al. Secure wireless communication of brain–computer interface and mind control of smart devices enabled by space-time-coding metasurface. *Nat Commun* **16**, (2025).
- Li L. et al. Machine-learning reprogrammable metasurface imager. *Nat Commun* **10**, (2019).
- Li, L. et al. Intelligent metasurface imager and recognizer. *Light Sci. Appl* **8**, 97 (2019).
- Ma Q., et al. Smart metasurface with self-adaptively reprogrammable functions. *Light Sci Appl* **8**, (2019).
- Liu, C. et al. A programmable diffractive deep neural network based on a digital-coding metasurface array. *Nat. Electron* **5**, 113–122 (2022).
- Qian C. et al. Dynamic recognition and mirage using neuro-metamaterials. *Nat Commun* **13**, (2022).
- Gu, Z., Ma, Q., Gao, X., You, J. W. & Cui, T. J. Direct electromagnetic information processing with planar diffractive neural network. *Sci. Adv.* **10**, eado3937 (2024).
- Gao, X. et al. Programmable surface plasmonic neural networks for microwave detection and processing. *Nat. Electron* **6**, 319–328 (2023).
- Gao, X. et al. Terahertz spoof plasmonic neural network for diffractive information recognition and processing. *Nat. Commun.* **15**, 6686 (2024).
- Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
- Wang, T. et al. Image sensing with multilayer nonlinear optical neural networks. *Nat. Photonics* **17**, 408–415 (2023).
- Yildirim M., Dinc N. U., Oguz I., Psaltis D., Moser C. Nonlinear processing with linear optics. *Nat Photonics*, (2024).
- Wanjura, C. C. & Marquardt, F. Fully nonlinear neuromorphic computing with linear wave scattering. *Nat Phys*, (2024).
- Pour Fard, M. M. et al. Experimental realization of arbitrary activation functions for optical neural networks. *Opt. Express* **28**, 12138–12148 (2020).
- Zuo Y. et al. All-optical neural network with nonlinear activation functions. *Optica* **6**, (2019).
- Ryou A. et al. Free-space optical neural network based on thermal atomic nonlinearity. *Photonics Res* **9**, (2021).
- Yan T. et al. Fourier-space Diffractive Deep Neural Network. *Phys Rev Lett* **123**, (2019).
- Huang Z. et al. Pre-sensor computing with compact multilayer optical neural network. *Sci Adv* **10**, (2024).
- Guo X., Barrett T. D., Wang Z. M., Lvovsky A. I. Backpropagation through nonlinear units for the all-optical training of neural networks. *Photonics Res* **9**, (2021).
- Ning Y. M., Ma Q., Xiao Q., Gu Z., Cui T. J. Reprogrammable Non-linear Transmission Controls Using an Information Metasurface. *Adv Opt Mater*, (2023).
- Zhou, T. et al. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **15**, 367–373 (2021).
- Ashtiani F. et al. A surface-normal photodetector as nonlinear activation function in diffractive optical neural networks. *APL Photonics* **8**, (2023).

38. Song A., Murty Kottapalli S. N., Goyal R., Schölkopf B., Fischer P. Low-power scalable multilayer optoelectronic neural networks enabled with incoherent light. *Nat Commun* **15**, (2024).
39. Natarajan P., Nevatia R. Online, Real-time Tracking and Recognition of Human Actions. In: *2008 IEEE Workshop on Motion and video Computing* (2008).
40. Harjanto F., Wang Z., Lu S., Feng D. D. Evaluating the impact of frame rate on video based human action recognition. In: *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*. Association for Computing Machinery (2012).
41. Tu Y. et al. *PlayerOne: Egocentric World Simulator*. arXiv (2025).
42. Singh S. P., Rahkonen T., Leinonen M. E., Parssinen A. A. 290 GHz Low Noise Amplifier Operating above $f_{\max}/2$ in 130 nm SiGe Technology for Sub-THz/THz Receivers. In: *2021 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)* (2021).
43. Hu L. et al. A 110-170 GHz Wideband LNA Design Using the InP Technology for Terahertz Communication Applications. *Micro-machines* **14**, (2023).
44. Singh, S. P., Rahkonen, T., Leinonen, M. E. & Parssinen, A. Design Aspects of Single-Ended and Differential SiGe Low-Noise Amplifiers Operating Above $f_{\max}/2$ in Sub-THz/THz Frequencies. *IEEE J. Solid-State Circuits* **58**, 2478–2488 (2023).
45. Wang, L., Shen, Y., Qian, Y., Ding, Y. & Hu, S. A 220-GHz LNA With 9.7-dB Noise Figure and 24.6-dB Gain in 40-nm Bulk CMOS. *IEEE Trans. Circuits Syst. II Express Briefs* **72**, 113–117 (2025).
46. Ji D. F., Niu B., Tao H. Q., Chen T. S., Wang W. B. A. D Band Zero Bias Detector Chip Using Schottky Diode. In: *2022 Photonics & Electromagnetics Research Symposium (PIERS)* (2022).
47. Ludwig F. et al. Graphene field-effect transistors as THz detectors: Distinguishing between resistive self-mixing and the hot-carrier thermoelectric effect. In: *2023 48th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)* (2023).

Acknowledgements

The work is supported by the National Natural Science Foundation of China (62301147, Q.M., 62288101, T.J.C., and 92167202, Q.M.), the National Key Research and Development Program of China (2022YFA1404903, Q.M.), Jiangsu joint laboratory of multidimensional perceptual information technology (bm2022017, T.J.C.), Special Fund for Key Basic Research in Jiangsu Province (BK20243015, T.J.C.), Natural Science Foundation of Jiangsu Province (BK20230822, Q.M.), the Major Project of Natural Science Foundation of Jiangsu Province (BK20212002, T.J.C., and BK20210209, T.J.C.), Young Elite Scientists Sponsorship Program by CAST (2022QNRC001, Q.M.), the State Key Laboratory of Millimeter Waves, Southeast University, China (K201924, T.J.C.), the Fundamental Research Funds for the Central Universities (2242023K5002, T.J.C., 2242018R30001, T.J.C., 2242022R20017, T.J.C.), the 111 Project (111-2-05, T.J.C.), and the China Postdoctoral Science Foundation (2021M700761, Q.M., 2022T150112, Q.M.).

Author contributions

T.J.C. and Q.M. initiated the plan and supervised the entire study. Q.M. and Y.M.N. conceived the idea of this work. Y.M.N. carried out the structural design and carried out the simulations. Y.M.N., Q.X., Z.G., Q.W.W., L.C., and R.S.L. carried out the measurements. Y.M.N. and Q.X. carried out the data analyses. Y.M.N., Q.M., and Q.X. wrote the manuscript. J.W.Y., X.X.G., and T.J.C. reviewed the manuscript. All authors discussed the theoretical aspects and numerical simulations, interpreted the results and reviewed the manuscript.

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/s41467-025-65275-0>.

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

Peer review information *Nature Communications* thanks Emir Salih Magden, who co-reviewed with Bahrem Serhat DanisFarshid Ashtiani; and Shuang Zhang for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025