AFRL-RX-WP-TR-2021-0222

SMART GLASS FOR OBJECT RECOGNITION

Zongfu Yu
University of Wisconsin-Madison

29 SEPTEMBER 2021
Final Report

AIR FORCE RESEARCH LABORATORY
MATERIALS AND MANUFACTURING DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433-7750
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RX-WP-TR-2021-0222, HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

BUSKOHL.PHILIP.REUBEN.1281844373 Digitally signed by BUSKOHL.PHILIP.REUBEN.1281844373 Date: 2022.01.04 11:03:13 -05'00'

PHILIP BUSKOHL
Work Unit Manager
Soft Matter Materials Branch
Functional Materials Division
Materials and Manufacturing Directorate

LEEVER.BENJAMIN.J.1262915170 Digitally signed by LEEVER.BENJAMIN.J.1262915170 Date: 2022.01.12 15:51:05 -05'00'

BENJAMIN LEEVER
Branch Chief
Soft Matter Materials Branch
Functional Materials Division
Materials and Manufacturing Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
29 September 2021	Final	22 June 2020 – 31 August 2021

4. TITLE AND SUBTITLE
Smart Glass for Object Recognition

5a. CONTRACT NUMBER
FA8650-20-1-7028

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Zongfu Yu

5d. PROJECT NUMBER
DARPA

5e. TASK NUMBER

5f. WORK UNIT NUMBER
X1T7

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Wisconsin System
Research & Sponsored Programs
Madison, WI 53715-1218

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory
Materials and Manufacturing Directorate
Wright-Patterson Air Force Base OH
45433-7750
Air Force Materiel Command
United States Air Force

Defense Advanced Research
Projects Agency
3701 North Fairfax Drive
Arlington VA 22203-1714

10. SPONSORING/MONITORING AGENCY ACRONYM(S)
AFRL/RXAS

11. SPONSORING/MONITORING AGENCY REPORT NUMBER(S)
AFRL-RX-WP-TR-2021-0222

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
This report contains color.

14. ABSTRACT *(Maximum 200 words)*

The goal of the project is to experimentally demonstrate a smart glass that can recognize objects. The glass does not consume energy and has no computer chips. It uses the physics of wave dynamics to perform analog artificial neural computing inside a nanostructured material. This smart glass can be as small as a few millimeters and last forever unless mechanically destroyed. The demonstration of a working prototype serves as the foundation to explore exciting possibilities in using wave dynamics for analog computing. The prototype will build on the concept of optical waves passing through a nanophotonic medium can perform artificial neural computing. Complex information is encoded in the wavefront of the input light. The medium transforms the wavefront to realize sophisticated computing tasks such as image recognition. At the output, the optical energy is concentrated in well-defined locations, which, for example, can be interpreted as the identity of the object in the image.

15. SUBJECT TERMS
Smart Glass, Optical Neural Network, Neuromorphic Metasurface, Resonator, Recurrent Neural Network

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT:	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON (Monitor)
a. REPORT	b. ABSTRACT	c. THIS PAGE			Philip Buskohl
Unclassified	Unclassified	Unclassified	SAR	33	19b. TELEPHONE NUMBER *(Include Area Code)* (937) 255-9152

Table of Contents

Section **Page**

LIST OF FIGURES

ACKNOWLEDGEMENTS

We acknowledge the support of the Defense Advanced Research Projects Agency.

1. INTRODUCTION

1.1 Smart Glass

Object recognition has been exploited in a wide range of applications, including image annotation, vehicle counting and tracking, pedestrian detection, and facial detection and recognition. Artificial neural networks (ANNs) which are multi-layer machine learning models have achieved significant success in various object recognition tasks. Although ANNs have become ubiquitous for applications on a daily basis, concomitant challenges emerge as the demand for computational speed and power efficiency increases rapidly with the explosion of data volume and the availability of mobile devices with computer vision features. Endeavors have focused on developing electronic architectures that are faster and more energy-efficient, but the hardware performance is still limited by the electronic clock rate and ohmic loss. An alternative approach is optical neural networks (ONNs), which utilize photonic elements and networks to construct a layered feed-forward architecture following the network structure of digital ANNs. In an ONN, the optical signals are manipulated by layers of elements in the sequence that perform transformations and weighting along the light propagation direction. The transformations and weights are pre-trained to enable the network to perform device-specific computation tasks. The computational speed of ONNs is primarily characterized by the speed of light propagating in the system and the passive nature of most optical operations introduces no additional power consumption after optical input is generated. Furthermore, digital ANNs are vulnerable to cyberattacks, while ONNs have computational algorithms hardcoded in the system so that security is guaranteed.

Most of the reported ONNs [1]–[3] follow the conventional layered architecture and utilize light propagation through the network. Here we propose a platform that compresses the series of transformations into a single layer of computational operation using metasurface. A metasurface[4], [5] is an engineered two-dimensional structure with millions of subwavelength scatterers called “meta-units” as its building block and can offer complete and precise control of the wavefront of light. Meta-units can scatter light with designed modulations to optical phase and amplitude depending on their morphology. The collective interference of scattered light from millions of meta-units results in an optical wavefront as designed. This technology has shown exciting potential in various applications such as flat lenses and planer holograms in the recent decade. For example, a meta-lens can condense the bulky compound lens system of conventional objects or camera lenses into a flat device and still minimize optical aberrations in the system. Furthermore, metasurfaces composed of birefringent meta-units with asymmetric cross-sectional shapes can modify optical wavefronts of incident light with different orthogonal polarization states in a complete and independent manner, a technique known as polarization multiplexing.

Our ONN is a transmissive metasurface, dubbed “smart glass”, with its 2D phase distribution trained for object recognition. We demonstrated object recognition using the smart glass with phase-only modulation. We also demonstrated a polarization-multiplexing smart glass that

performs distinct recognition tasks using light linearly polarized in two orthogonal directions. This smart glass alone acts as a passive computer that operates at the speed of light and without consuming any energy. The compression of multiple computational layers into a singlet largely reduces the number of weight parameters to train. The subwavelength dimensions of meta-units enable efficient parallel computing of high-resolution data. Moreover, physics-based computing that relies on intrinsic and engineered material properties could offer security beyond digital encryption.

1.2 Resonant recurrent neural network

Larger and deeper neural networks require new and high-efficiency computing architectures. The search for alternative computing has recently intensified. One interesting candidate is to use physical waves such as light to perform analog computing. Such analogy computing enjoys the benefits of intrinsic parallelism, and it can be extremely energy efficient. The great promise also comes with great challenges. The very fact that gives rise to the lightning speed of wave computing also leads to one major challenge: the transient nature of propagating waves makes it difficult to construct a memory in the wave domain. Since memory is indispensable for computing temporal data, today researchers must resort to other means to realize the effect of memory such as optoelectronic conversion, routing through long waveguides, and random internal feedback.

As a natural memory, resonance is well suited as the building block for scalable recurrent neural networks (RNNs). RNNs have been widely used for speech recognition[6] and synthesis[7], and machine translation[8]. In digital RNNs, memory has been realized by constructing digital feedbacks [9], and the same feedback can be realized by wave reflection at the boundaries of resonators. Different types of analog RNNs have been previously investigated. One strategy is to faithfully follow digital feedbacks in RNNs by routing the output of a computing device back to its input. For example, one can feed the output of optical networks back to the input via long optical waveguides or electro-optical conversion. By merely functioning as a mirror copy of digital RNNs, this strategy does not take advantage of intrinsic feedback mechanisms in wave physics. Recently,[3]. proposed to use the intrinsic feedback in disorder structures to increase the expressive power and reduce the size of analog neural computing. (Hughes et al., 2019)showed that such internal feedback mechanism can be used to realize RNNs. For a long time, reservoir computing in the optical domain has also explored complex temporal dynamics to realize the memory effect. In all these works, the memory is implicitly built into the complex structures. One cannot easily identify the spatial distribution or the temporal characteristics of memory. While in principle these analog systems can perform certain temporal computation, physical intuition and interpretation are lacking. RNNs have the vanishing gradient problem in nature: the gradient of the loss function decays exponentially with time. As a result, it is difficult for vanilla RNNs to keep track of arbitrary long-term dependencies in the input sequences. Long short-term memory (LSTM) is designed to solve the problem of learning long-term temporal dependencies. It incorporates cell states and gates such that events from the remote past can have current impacts. It is difficult to construct and interpret LSTM in existing wave computing systems because their temporal characteristics are

buried in a black box. However, in the resonance system, we can include resonators with different lifetimes to realize short-term and long-term memory. This explicit form of memory makes it easy to construct scalable RNNs and advanced recurrent models such as LSTM. As the quest for the new computing intensifies, resonance could play an important role in emerging wave-based analog computing in general.

We demonstrate that resonance can be used to construct stable and scalable recurrent neural networks. By including resonators with different lifetimes, the computing system develops both short-term and long-term memory simultaneously.

2 METHODS, ASSUMPTIONS, AND PROCEDURES

2.1 Smart Glass

2.1.1 ONNs for object recognition

Figure 2.1 shows a schematic that illustrates the configuration of the ONN system in this study. It is composed of three components: an input, the smart glass, and an output. The input comprises an object to be recognized (e.g., a digit, a letter) and is placed at a distance (object distance) in front of the smart glass. The far-field diffraction pattern of the input is incident upon the smart glass and exits with a modulated wavefront. The modulated light then propagates over another distance (imaging distance) and arrives at the output layer, producing a diffraction pattern. Object recognition is accomplished by concentrating optical intensity to a specific location on the output plane, depending on the classification label of the input object.

The ONN system is designed for near-infrared light at λ=1,550 nm. The input objects are binary images of handwritten numerical digits in the MNIST dataset, or typed letters in several fonts, with a size of 500 $\lambda \times$ 500 λ (**Figure 2.1**). The object and imaging distances where light propagates in the air (n=1) are both 2000 λ. The smart glass is composed of a single layer of metasurface on a glass substrate, which has a thickness of 322.58 λ (500 μm) and a refractive index of 1.44. The metasurface is patterned on the backside of the substrate and is circular with a radius of 250 λ. A

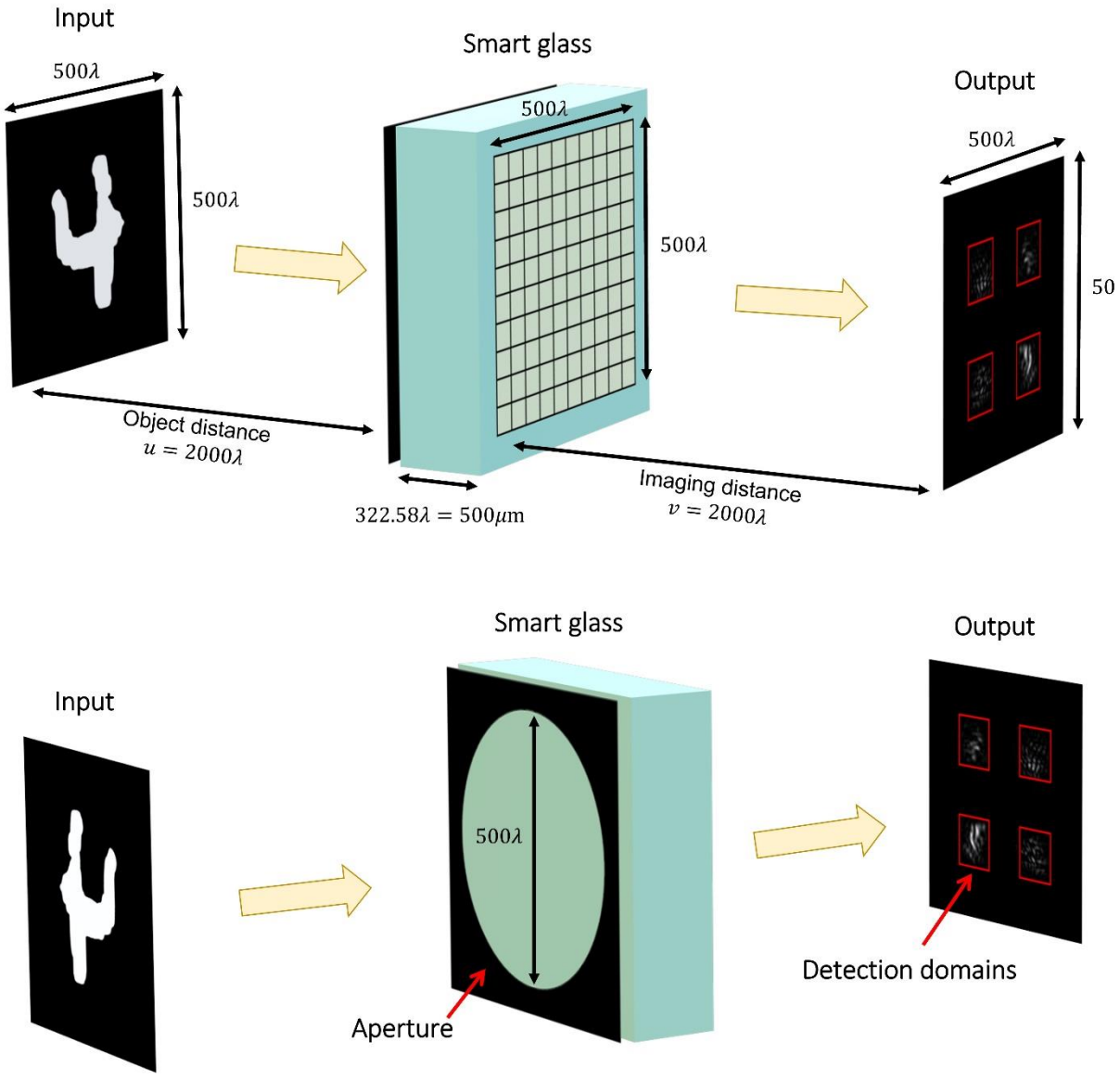

Figure 2.1 **Operation of a smart glass for object recognition.** Schematics illustrating that a smart glass channels light from an input image preferentially onto one of several detection domains in the output plane. The yellow arrows indicate the direction of light propagation. The red squares in the output plane define the detection domains. Note that in experiments the metasurface smart glass is defined on the back surface of a transparent substrate and the aperture is placed on the front surface of the substrate.

circular aperture with the same size as the metasurface is positioned on the front surface of the substrate to block stray light in the experiment (**Figure 2.1**).

An ANN simulating the above system on a digital computer is used for training the smart glass. In the training process, both the object and the metasurface are digitalized into 1000 × 1000 pixels, with a spatial resolution of 750 nm. The input image can be considered as a collection of uniformly

distributed dipoles, all assumed to lay in the same direction parallel to the input plane. The propagation of the image for a distance of d can be calculated as the sum of dipole radiations, given by

$$E(\boldsymbol{r_j}, d) = \sum_i E(\boldsymbol{p_i}, \boldsymbol{r_i}, \boldsymbol{r_j}, d)$$

$$= \sum_i \frac{1}{4\pi\epsilon_0}\left\{k^2(\boldsymbol{n_{ij}} \times \boldsymbol{p_i}) \times \boldsymbol{n_{ij}}\frac{e^{ikr}}{r} + \left[3\boldsymbol{n_{ij}}(\boldsymbol{n_{ij}} \cdot \boldsymbol{p_i}) - \boldsymbol{p_i}\right]\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)e^{ikr}\right\}, \qquad \text{Eq 2.1}$$

where i and j denote the i^{th} and j^{th} locations in the input and output planes, respectively; $\boldsymbol{r}$ is the coordinate of a location; $\boldsymbol{p}$ is the electric dipole moment; $\boldsymbol{n_{ij}}$ is the unit vector in the direction of $\boldsymbol{r_j} - \boldsymbol{r_i}$; ϵ_0 is the permittivity of free space; $k = 2\pi/\lambda$ is the wavenumber. For a binary image, the magnitude of dipole moment $\boldsymbol{p}$ is 1 in the digit or letter, and 0 in the background.

To accelerate computation, a 3FFT method is implemented following the Huygens-Fresnel principle. Consider waves propagate from a plane to another, separated by a distance d, the relation between the phasor distributions of the wave on the two planes is given by

$$u(d) = u(0) \otimes k \qquad \text{Eq 2.2}$$

where $u(0)$ and $u(d)$ are the phasor profiles on the two planes, and k is the propagator kernel of the aspherical wave modulated by an inclination factor. By applying Fourier transform on both sides, the relation becomes

$$U(d) = U(0) \cdot K \qquad \text{Eq 2.3}$$

here the upper case denotes the Fourier transforms of each term. The phasor profile $u(d)$ can be quickly obtained by numerically calculating the product of the Fast-Fourier-transforms (FFTs) of $u(0)$ and k, followed by calculating inverse FFT of the product, to avoid the computationally inefficient summation or convolution over pixels of the planes.

In this manner, the binary optical input is first propagated over the object distance, to the metasurface plane. Modulations are subsequently applied to the calculated fields on the metasurface plane (**Figure 2.1**). In this study, we only design the metasurface based on its modulation of the optical phase. The modulated field propagates for the image distance and the resulting output field is used to evaluate the loss function.

In the experiment, the output optical intensity is captured by cameras or photodetectors. Detection domains are designated on the output plane representing the classification labels (**Figure 2.1**). The loss function evaluates the cross-entropy between the measured intensity distribution and the target intensity distribution, given by

$$Loss\ function = \sum_i p_i \cdot \log(q_i)\,, \qquad \text{Eq 2.4}$$

where q_i is the proportion of intensity in the i^{th} domain to the total intensity and p_i is the target value, which is 1 if the domain matches with the label of the input and 0 if not. The phase responses of the 1000×1000 pixels of the metasurface smart glass are trained to minimize the loss function using the "Adam" optimization algorithm adapted from the stochastic gradient-based optimization method.

Measures are taken to increase the robustness of the trained ONN against experimental errors. For example, to increase the tolerance to minor mispositioning of the optical components, the detection domains are designed to have a separation between each other (**Figure 2.1)**). An auxiliary loss term, given by the proportion of intensity in the detection domains to the total intensity, is subtracted from the loss function, to increase the contrast between the domains of interest and

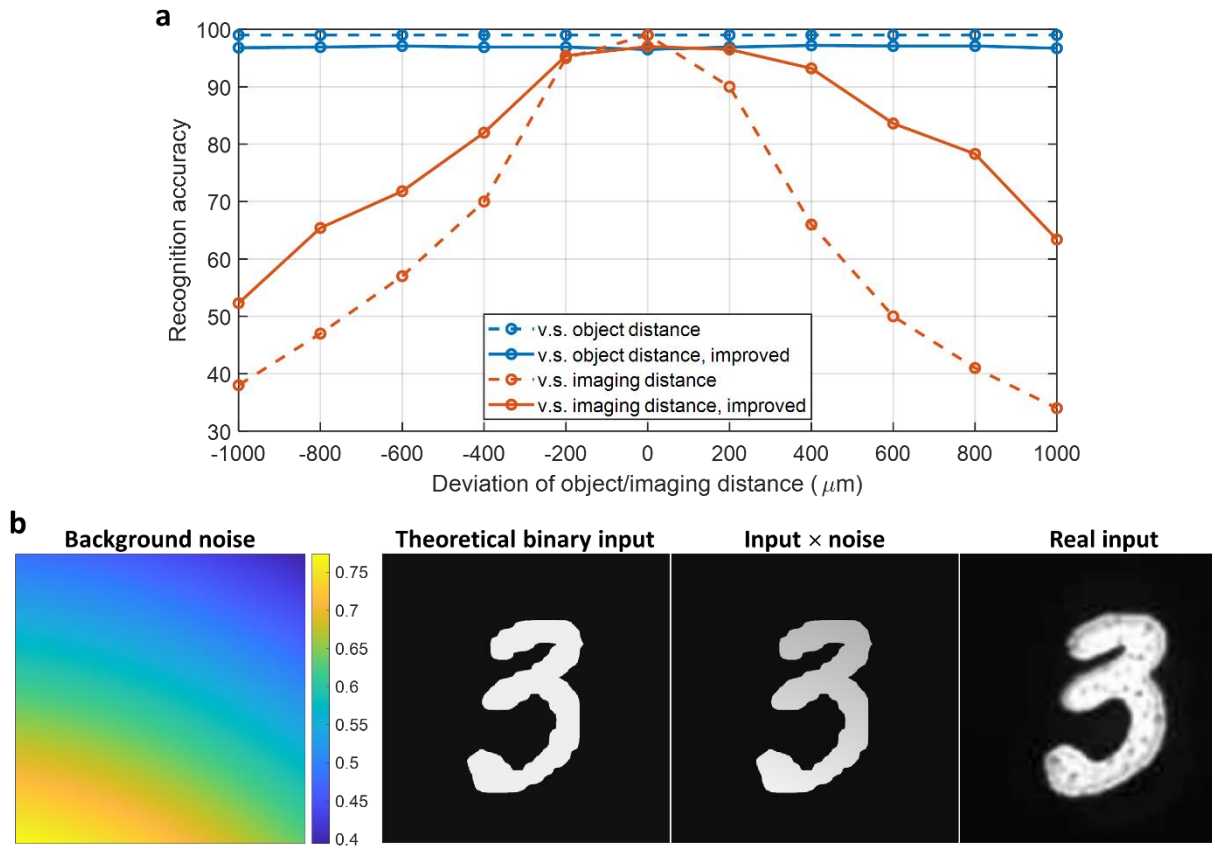

Figure 2.2 **Measures to increase robustness of smart glasses against experimental errors.** (a) Simulated recognition accuracies of a smart glass as a function of the deviation of object/imaging distances, showing that the robustness of the smart glass is improved by taking the measures described in the text. Solid lines: accuracies with the measures taken. Dashed lines: accuracies without taking any measure. (b) An example of introducing a nonuniform background ('noise') during the training of the smart glass. From left to right: a randomly generated nonuniform background, a binary input image, the product of input and 'noise' used in the training process, a real input in the experiment.

background via the minimization of the loss function. Besides, the input images are randomly displaced along with their in-plane directions by a maximum of 10% of the image size in the training process to ensure the tolerance of the trained ONN to the error in input positions. By taking the above measures, the trained ONN also shows robustness against the errors in the positions of the components along the optical axis. **Figure 2.2 (a)** shows that with the smart glass fixed, the recognition accuracy is highly robust against the object distance. Although more sensitive to the imaging distance, the accuracy remains high when the imaging distance deviates by up to 15%. Furthermore, a spatially slowly varying background noise, with a gaussian intensity profile and its size, position, and magnitude randomly designated, is applied to the input images that are used in the training, to take account of the experimental non-uniformity of input optical intensity. **Figure 2.2 (b)** depicts a specific example of such a background intensity variation.

2.1.2 Design and Fabrication of metasurface

The metasurface is made of amorphous silicon for its transparency in the near-infrared and is composed of meta-units situated in a square lattice with a periodicity of 750 nm. The height of the meta-units is 1 μm. A library of meta-units is generated with individual phase responses

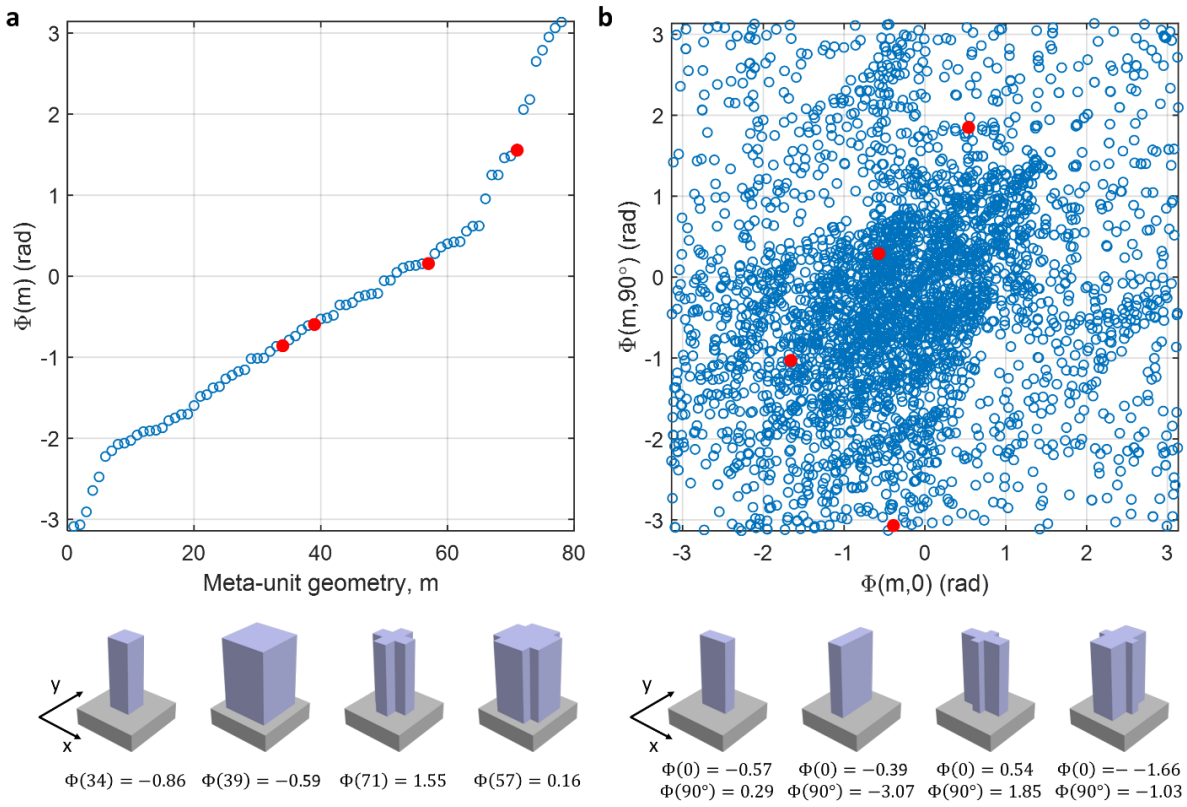

Figure 2.3 **Isotropic and birefringent meta-unit libraries.** (a) Phase responses, $\Phi(m)$, of meta-units in the polarization-independent library, where *m* denotes the index of the meta-unit geometry. (b) Phase response, $\Phi(m,\theta)$, of meta-units in the birefringent library, where θ denotes polarization direction measured from the x axis. Each blue circle represents one meta-unit. Red solid circles represent a few examples of meta-units illustrated below.

determined by Rigorous Coupled Wave Analysis (RCWA). Here, the meta-units have two cross-sectional archetypes, square and cross-shaped; their phase responses can be either polarization-independent or birefringent. In the latter case, the meta-unit pillars have a non-unity aspect ratio in the cross-section, and such meta-units are used for polarization-multiplexing smart glasses. **Figure 2.3 (a)** and **(b)** show the calculated phase responses, $\Phi(m,\theta)$ (m denotes the meta-unit geometry and θ is the polarization direction of light), of the meta-units used in the polarization-independent and birefringent libraries, respectively.

Note that the library is discrete in its phase responses because the minimal increment of meta-unit dimensions is limited by the fabrication resolution of electron-beam lithography. We use a customary error-minimization algorithm that sweeps the entire parameter space of the library to find the optimal meta-unit for each position of the meta-surface. The figure of merit used is the average phase error over the entire meta-surface lattice. Suppose at position (x, y) we have the desired phase response for a certain polarization, $\Phi(x,y,\theta)$. For each position, the meta-unit with a phase response that deviates the least from the desired phase response is found. By sweeping over all the lattice positions, we can generate an optimal meta-surface layout with a collective phase profile that replicates the desired profile of phase modulation.

In the fabrication process, amorphous silicon films of thickness 1 μm are deposited by plasma-enhanced chemical vapor deposition on 500-μm thick fused quartz substrates. An electron beam lithography system (Elionix ELS-G100) is used to define the metasurface pattern on a double-layer electron-beam Resist (PMMA 495k A4 and 950A2) with a dose of 770 μC/cm^2 at a current of 2 nA. A 20-nm thick layer of E-Spacer (DisCharge H2Ox2) is spun on top of the double-layer

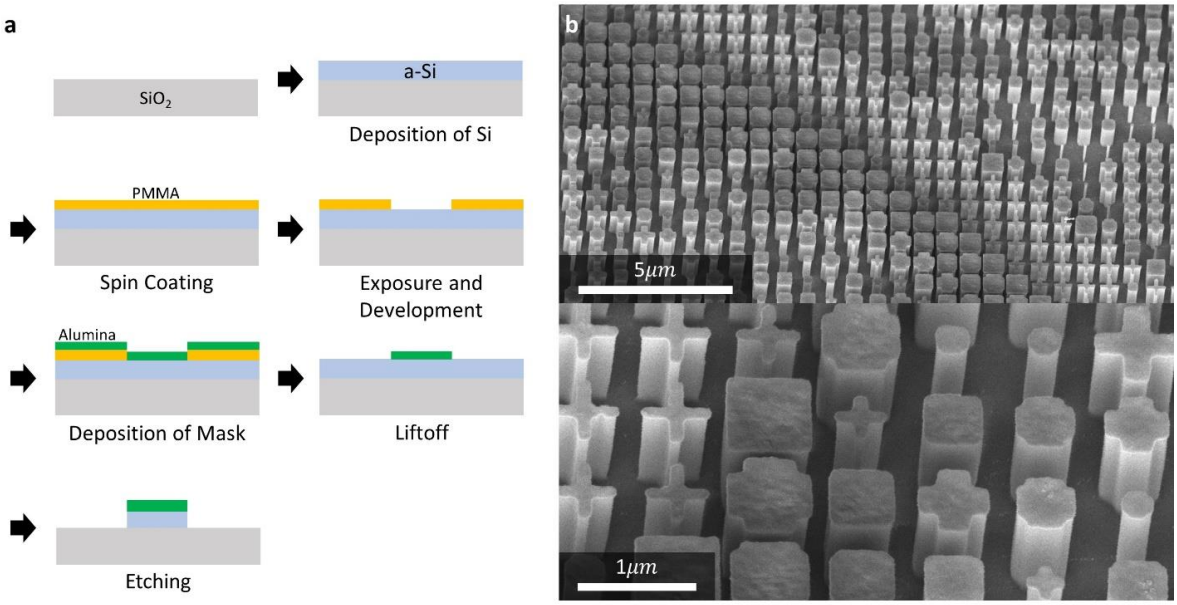

Figure 2.4 **Fabrication of metasurface smart glasses.** (a) Schematic illustrating the fabrication process. (b) SEM images showing meta-units with a variety of geometries used in a metasurface smart glass.

resist to alleviate the electron charging effect during exposure. The exposed resist is subsequently developed in a solution of 3:1 IPA: DI Water for 2 minutes. After the development of the resist, a 25-nm thick aluminum oxide layer is deposited on the top of the developed sample as an etching mask using electron-beam evaporation. Lift-off is performed by dissolving the remaining resist in acetone overnight and the meta-surface pattern is transferred from the aluminum oxide mask to the amorphous silicon film in an inductively coupled plasma (ICP) etcher (Oxford PlasmaPro 100 Cobra), with a homebrew etching recipe. The aluminum oxide mask is left on the devices because its thinness and dielectric nature make the optical impact of its presence negligible. Removal could be achieved by dissolving it in ammonium hydroxide without affecting the silicon or fused silica wafer. **Figure 2.4 (a)** summarizes the above fabrication process. A Scanning electron microscope (SEM) image shows an array of meta-units of a fabricated smart glass in **Figure 2.4 (b)**.

2.1.3 Experimental configuration

The experimental setup following the scheme of ANN is shown in **Figure 2.5 (a)**. A telecom laser beam (λ = 1,550 nm) is incident on a photomask to create an input optical object. The photomask contains a 2D array of objects (numerical digits or letters) that are transparent within the object and opaque outside of it. The photomask is made of emulsion photo-plotted on a mylar (Bo-PET, Biaxially-oriented polyethylene terephthalate) sheet (**Figure 2.5 (b)**). The incident beam has a diameter of ~3 mm, which is much larger than the size of individual input objects (0.775 mm × 0.775 mm) to provide relatively uniform illumination. The input image is relayed by a telescope with a unity magnification and the relayed image is superimposed on a square aperture so that stray light from adjacent object images is blocked (**Figure 2.5 (b)**). The optical wave from a single object is thus allowed to propagate over the object's distance. The diffraction pattern of the object is filtered by the circular aperture (made of aluminum foil, **Figure 2.1**and **Figure 2.5 (c)**) and then processed by the meta-surface smart glass. The output image is collected by a microscope with an objective focused on the output plane and measured by an InGaAs near-infrared camera (Princeton Instrument NIRvana 640). The optical intensities in the detection domains are extracted from the image and the input object is categorized according to the domain receiving the highest intensity. To test the polarization-multiplexing devices, a linear polarizer is inserted in front of the object mask so that an object is tested twice, one with horizontally polarized incident light and the other with vertically polarized incident light.

The experimental characterization of the ONN is conducted with at least 10 objects for each classification label to estimate the accuracy of object recognition. The tested objects are randomly chosen from a dataset that is not used in the training.

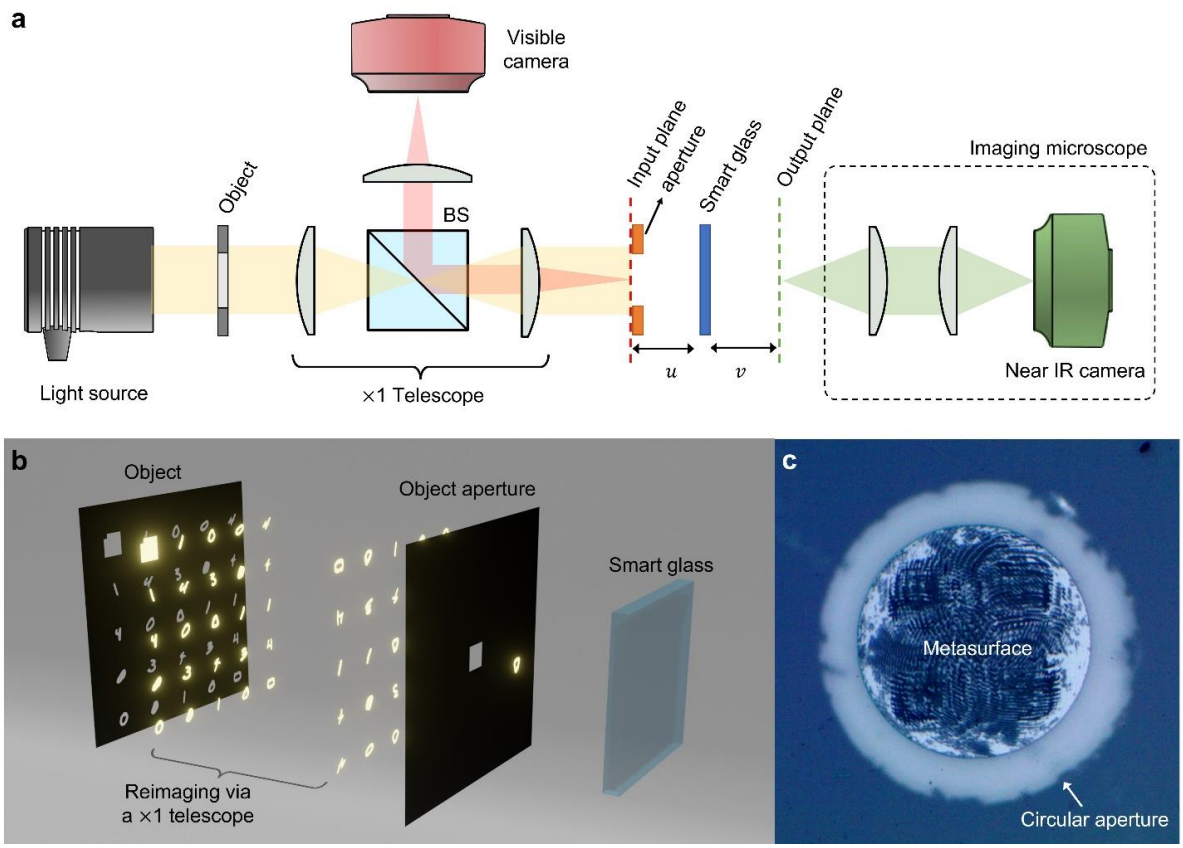

Figure 2.5 **Experimental configurations for testing the smart glass.** (a) Schematic of the experimental setup. BS: beam splitter. (b) An object plate contains an array of objects to be recognized (apertures defined on an opaque photomask). A square object aperture is placed on the input plane as shown in (a). The objects are reimaged on the input plane by a ×1 telescope and filtered by the aperture to allow light from a single object to propagate to the smart glass. (c) Optical microscopic image of a metasurface smart glass defined on the back surface of a glass substrate. A circular aperture (a ~ 1mm hole in an aluminum foil) is placed on the front surface of the substrate to reduce stray light.

2.2 Resonant Recurrent Neural network

We demonstrate the mathematical equivalence between digital RNN and resonant system. A digital RNN consists of many artificial neurons with memories. One neuron often connects to many others. Similarly, one resonator can couple to many, providing a scalable way to construct a large-scale analog computing system with memory. The coupling between resonators can be mediated through free space or waveguides. The coupling coefficients (e.g., connection weights) determine the function of computing. They will be trained in a similar way that neural networks

are trained. The trained coefficients can be physically implemented, for example, by adjusting the distance between resonators.

Next, we demonstrate a specific example where we train acoustic resonators to recognize the vowels spoken by a human. All computing is in the acoustic domain. The computing device [Fig. 10(a)] contains three parallel acoustic waveguides that couple to two rows of whispering gallery resonators **Figure 2.6**. Each row contains 60 resonators. The coupling between resonators is mediated through the waveguides. Depending on the identity of the sound, the computing device will route the energy of three different vowel sounds to three corresponding output waveguides. The input to the acoustic system is the replay of the audio recording of 10 vowel classes from 45 different male speakers and 48 different female speakers. We use a subset of the original in Ref **Figure 2.6**, which includes 279 recordings corresponding to three vowel classes, represented by ae, ei, and iy. The training is performed by using 80% of these recordings while the rest 20% is used to test the trained system. In the training, we can adjust geometric parameters, including the resonator positions l, the resonator sizes r, and the distances between resonators and waveguides

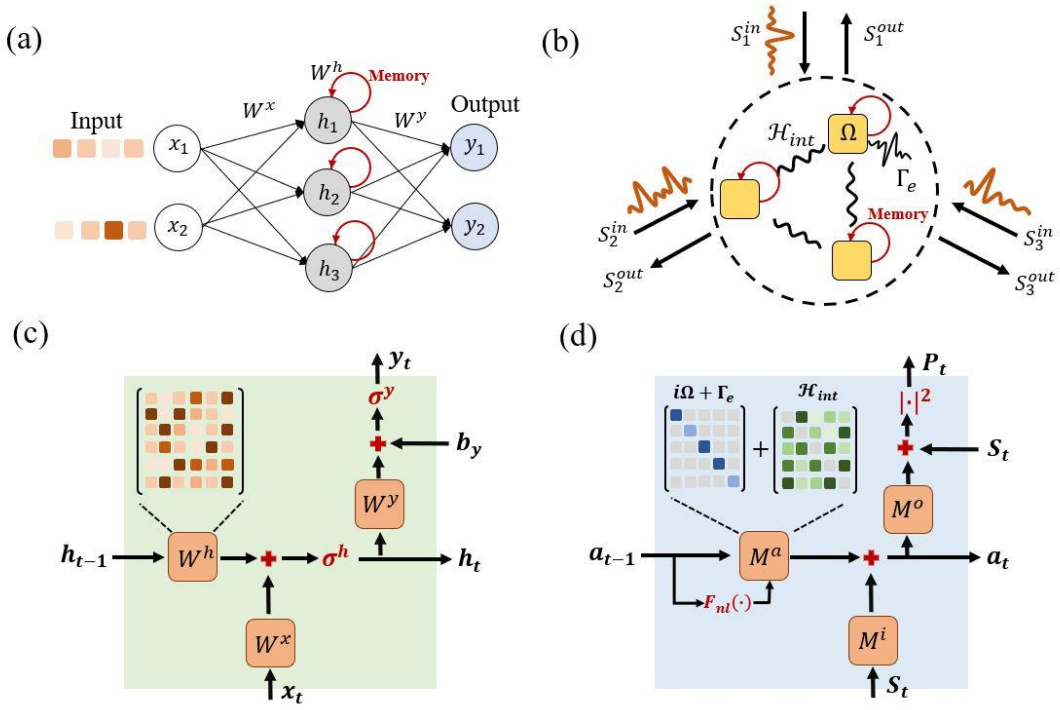

Figure 2.6 **Conceptual equivalence between of a RNN and a resonance system.** (a) General RNN architecture consisting of an input layer, hidden states with feedback loops and an output layer. The RNN can operate on discrete-time sequence. (b) Layout of a resonant recurrent network that can operate on continuous-time signal. Ω, Γ_e, and $\mathcal{H}_{int}$ represent resonance frequency, decay rate into channels, and coupling between resonators, respectively. (c) Diagram of a RNN cell composed of trainable parameters W^x, W^h, and W^y. (d) Diagram of the recurrence relation for the resonance system composed of trainable parameters M^i, M^a, and M^o.

d. The training goal is to minimize a loss function that is defined by the output of the three waveguides and ground-truth labels.

3 RESULTS AND DISCUSSION

3.1 Smart Glass

3.1.1 Recognition of 4 numerical digits

The prototype of smart glass (with a phase modulation shown in **Figure 3.1(a)**) is to recognize 4 numerical digits: {0, 1, 3, 4} from the MNIST[11] handwritten digit database. The smart glass is trained to concentrate light scattered from the binary image of one digit into one of the four square domains on the output plane as depicted in **Figure 3.1 (a)**. The measurements of 116 input digits (4 classes and N>25 for each class) result in recognition accuracy of 0.9828, while the training reports an accuracy reaching 0.9914. **Figure 3.1 (b)** shows a digit "4" on the input plane and the resulting intensity distribution on the output plane. **Figure 3.1 (c)** shows the confusion matrix summarizing the recognition results.

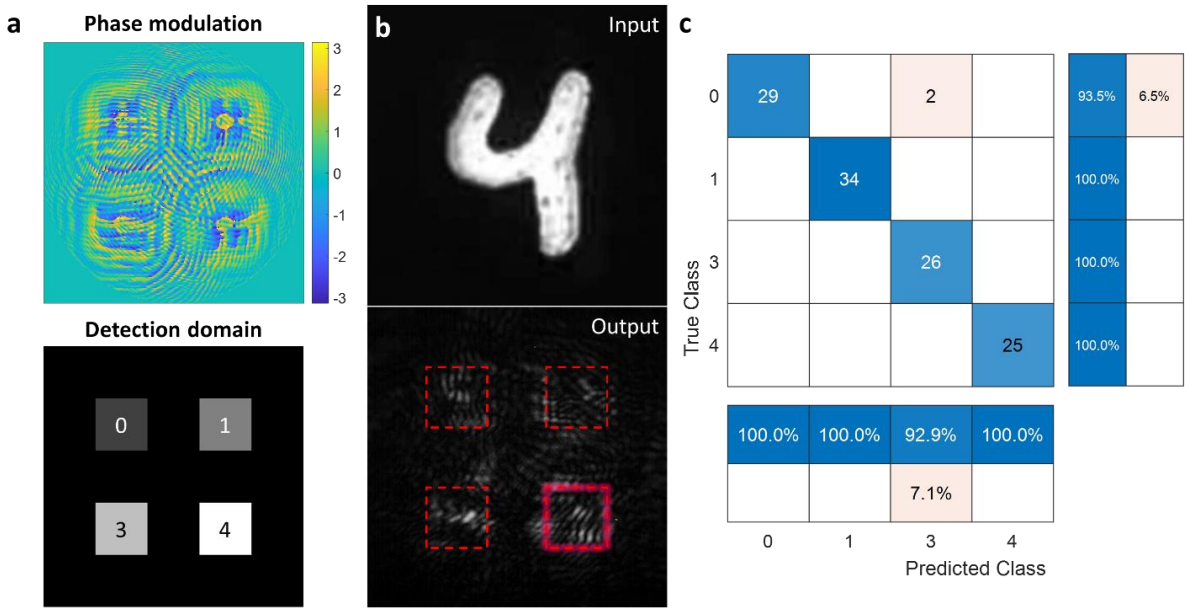

Figure 3.1 **Recognition of 4 types of handwritten numerical digits.** (a) Trained phase modulation on the metasurface and arrangement of detection domains on the output layer. (b) An example showing the recognition of a handwritten digit '4'. Upper panel: Input image. Lower panel: Intensity distribution on the output plane showing that the lower right detection domain (the one with glow) has the highest integrated optical intensity. (c) Confusion matrix summarizing the result of recognizing116 handwritten '0', '1', '3' and '4'. Each row on the matrix shows the instances in a true class and each column shows the instances in a predicted class. The lower table reports the proportions of correct (upper row) and incorrect (lower row) recognition for each predicted class, while the right table reports the proportions of correct (left column) and incorrect (right column) recognition for each true class.

3.1.2 Recognition of 10 numerical digits

We expanded the number of object classes to 10 digits, from 0 to 9, to further explore the capacity of our smart glass in object recognition. The trained optical phase profile of the meta-surface and 10 circular detection domains in a circular array are depicted in **Figure 3.2 (a)**. The measurements of 208 input digits (10 classes and N>10 for each class) result in recognition accuracy of 0.7837, while the training reports an accuracy reaching 0.865. Though much lower than that of the 4-digit

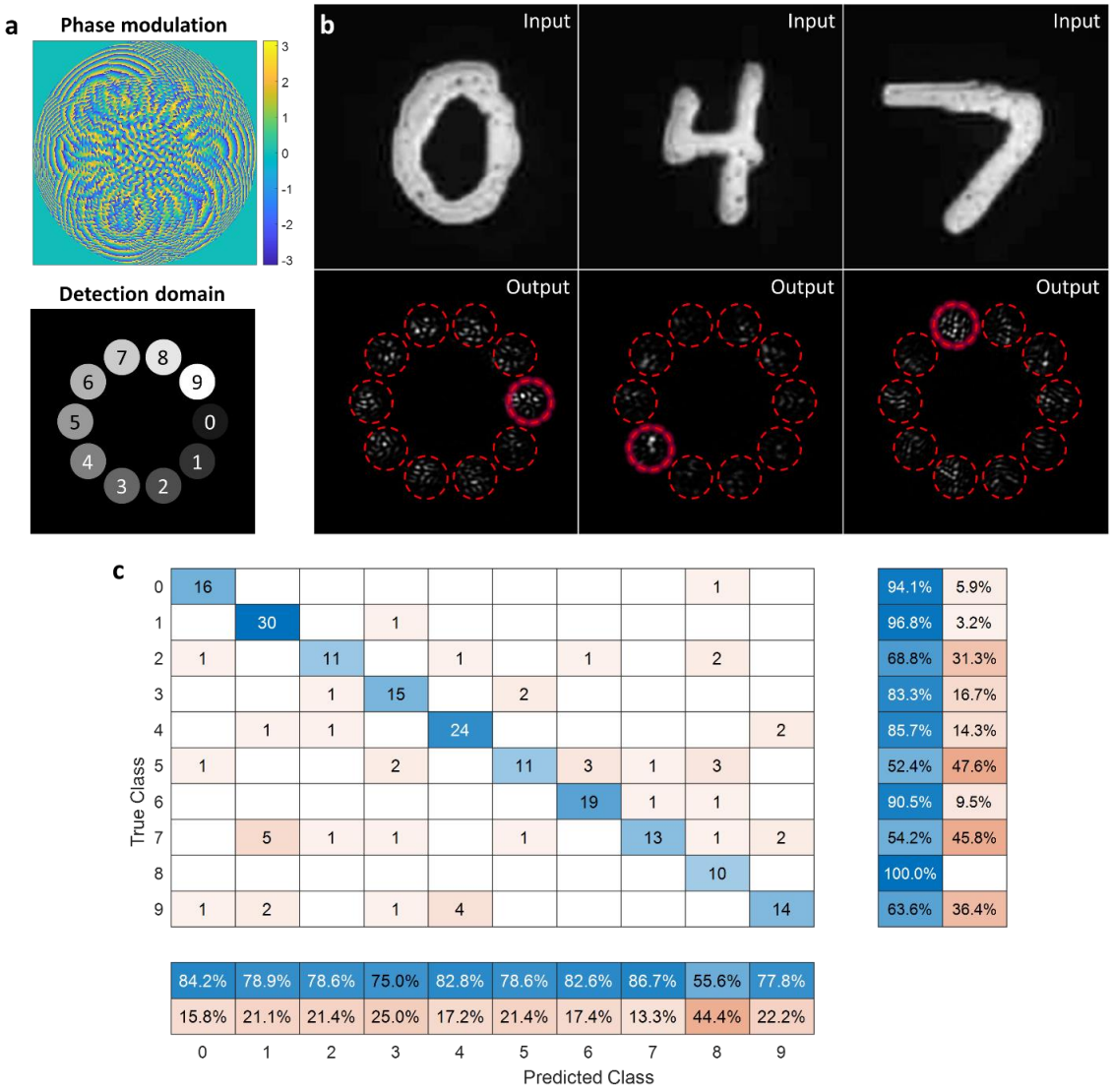

Figure 3.2 **Recognition of all 10 types of handwritten numerical digits.** (a) Trained phase modulation on the metasurface and arrangement of detection domains on the output layer. (b) Three examples showing the recognition of handwritten digits '0', '4' and '7'. Upper panels: Input images. Lower panels: Intensity distributions on the output plane showing that the detection domain corresponding to the identity of a digit has the highest integrated optical intensity. (c) Confusion matrix summarizing the result of recognizing 208 handwritten digits.

recognition device, the accuracy is still satisfactory. **Figure 3.2 (c)** shows the confusion matrix summarizing the recognition results.

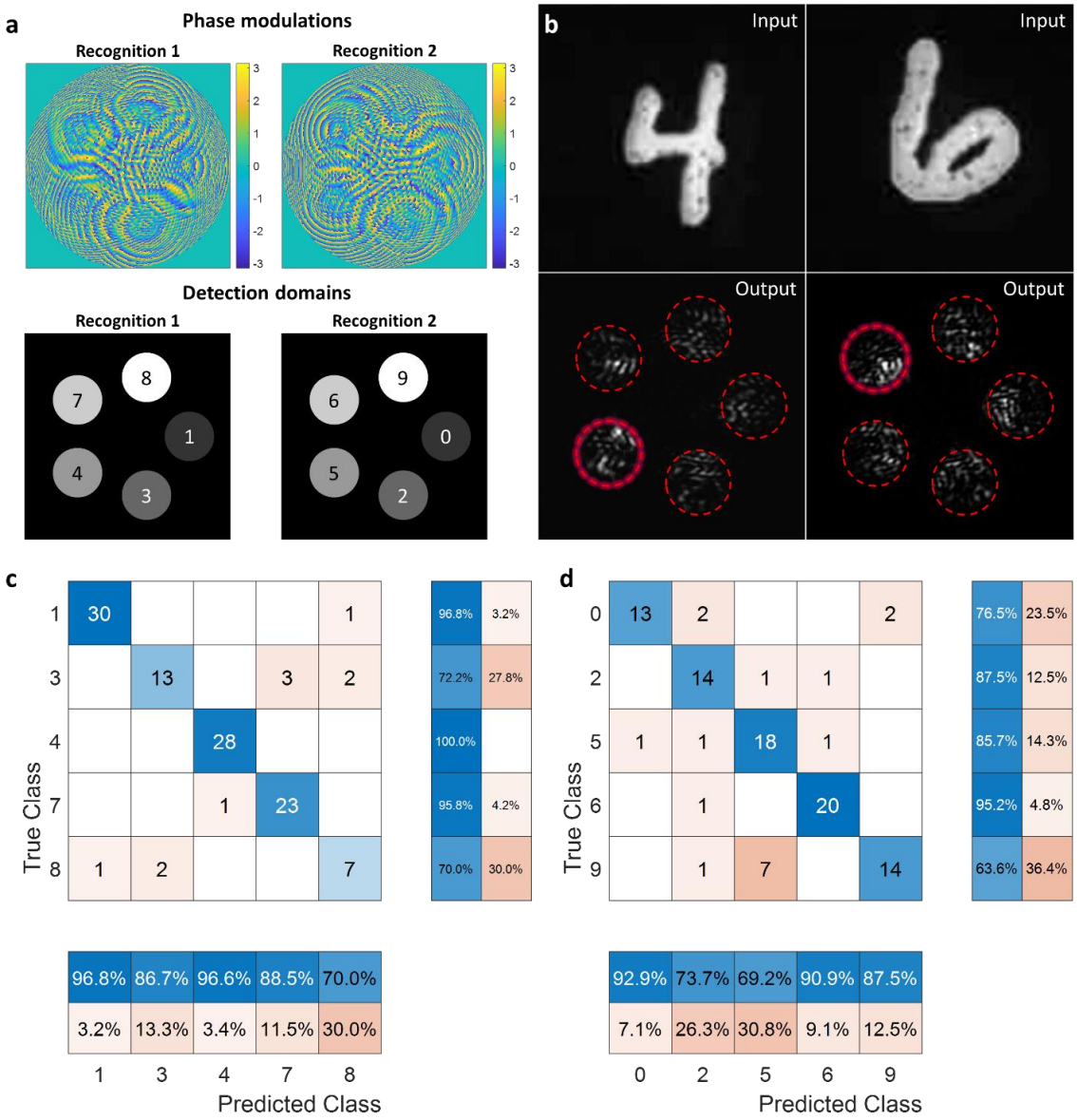

Figure 3.3 **Recognition of all 10 handwritten numerical digits using a polarization-multiplexing smart glass.** (a) Trained phase modulations at two orthogonal incident polarizations on the metasurface for recognizing two groups of digits: {1, 3, 4, 7, 8} and {0, 2, 5, 6, 9} and arrangement of detection domains on the output layer. (b) Two examples showing the recognition of handwritten digits '4' and '6'. Upper panels: Input images. Lower panels: Intensity distributions on the output plane showing that the detection domain corresponding to the identity of a digit has the highest integrated optical intensity. (c) and (d) Confusion matrices summarizing the result of recognizing the two groups of digits with 111 and 97 handwritten digits, respectively.

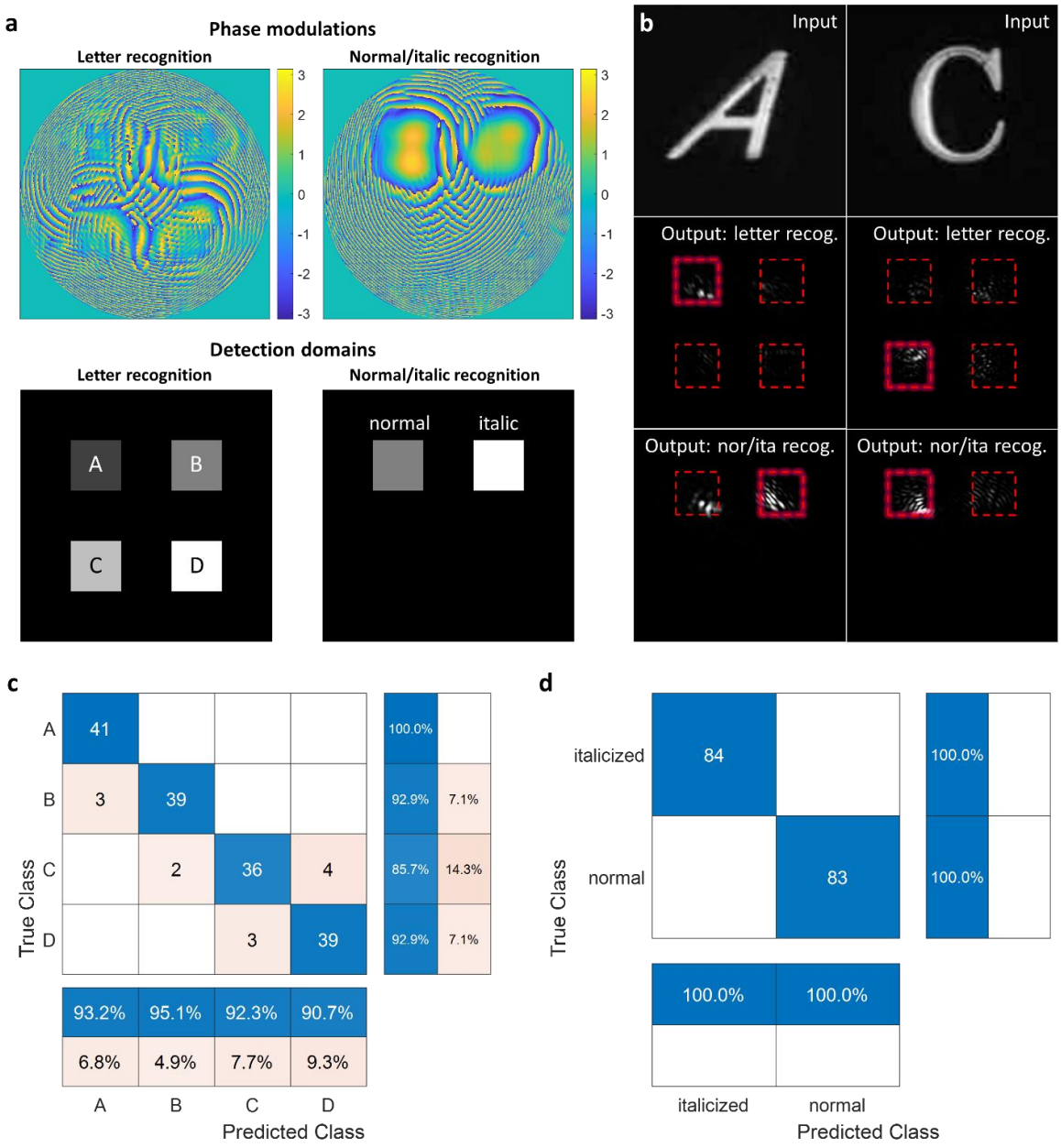

Figure 3.4 **Recognition of the identity and typographic styles of 4 letters using a polarization-multiplexing smart glass.** (a) Trained phase modulations at two orthogonal incident polarizations on the metasurface and arrangement of detection domains on the output layer. (b) Two examples showing the recognition of an italicized 'A' and a normal 'C'. Upper panels: Input images. Lower panels: Intensity distributions on the output plane showing that the detection domains corresponding to the identity and typographic style of a letter have the highest integrated optical intensity. (c) and (d) Confusion matrices summarizing the results of recognizing 168 letters.

3.1.3 Recognition of 10 numerical digits using polarization multiplexing

The 10-digit recognition is computationally a much more expensive task than categorizing only 4 digits. As such, we devise a polarization multiplexing strategy to reduce the complexity of the task. The 10 numerical digits are divided into two groups and their recognition is performed using light polarized in two orthogonal directions: the horizontally polarized light for the recognition of digits: {1, 3, 4, 7, 8} and the vertically polarized light for the recognition of digits: {0, 2, 5, 6, 9}. The smart glass is constructed using the birefringent meta-unit library to provide distinct phase modulations for light polarized in orthogonal directions (**Figure 3.3 (a)**). Five detection domains for recognition of each group of five digits are defined on the output plane as shown in **Figure 3.3 (a)**. The measurements of 111 and 97 input digits result in recognition accuracies of 0.9099 and 0.8144, for the two groups of digits, respectively, while the training reports accuracies reaching 0.9820 and 0.9691, respectively. **Figure 3.3 (c)** and **(d)** show the confusion matrices summarizing the recognition results. Note that the coverage of the optical phase provided by the birefringent meta-unit library is more discrete than that of the polarization-independent meta-unit library so that the phase responses of the fabricated birefringent meta-surface deviate from the desired phase profiles more than the non-birefringent devices. This issue can be addressed by including more archetypes of meta-units in the library.

3.1.4 Multiplexing & multitasking smart glass

The smart glass is designed to accomplish two distinct tasks using light with orthogonal polarizations: the light polarized in one direction is used for the recognition of typed alphabet letters and the light polarized in the orthogonal direction is for the recognition of the typographic types of the letters, i.e., if a letter is normal or italic. Four square detection domains are defined on the output plane, corresponding to 4 letters: {A, B, C, D} (**Figure 3.4 (a)**), and the two domains in the upper row are used for recognizing if a letter is normal or italic. Limited by the number of fonts available, the training and testing processes used the same dataset of 21 fonts so the accuracies of the two types of recognitions are both 1 reported by the training. The measurements of 168 letter inputs (N=21 for each letter and N=84 for each typographic type) show accuracies of 0.9281 and 1 for the letter recognition and typographic type recognition, respectively. **Figure 3.4 (c)** and **(d)** show the confusion matrices summarizing the recognition results.

3.1.5 Metasurface doublet for human face verication

We designed a double layered metasurface stack (“doublet”) and theoretically demonstrated human face verication with high accuracy. This metasurface doublet can compare two distinct gray-scale images of human faces (light wave transmitted through the gray-scale photos) and verify whether the images represent the same person or not. To realize this function, the doublet maps the input light from an image into an array on the output plane (i.e., an intensity pattern on 9 predefined regions on the camera plane) [**Figure 3.5 (a)**], and the similarity between two images is evaluated by calculating the Euclidean distance between two arrays corresponding to the two images (i.e., if the Euclidean distance is smaller than a threshold, there is a match between the two images) [**Figure 3.5 (c)**].

We used a dataset of photos of 100 persons, each person with 14 distinct photos. The data of 90 persons were used to train the metasurface doublet and the rest photos belonging to 10 persons were used as test. The result shows that when the threshold Euclidean distance is chosen to be ~0.8, we can have a small false acceptance rate of ~10% and a small false rejection rate of 10% [**Figure 3.5 (d)**].

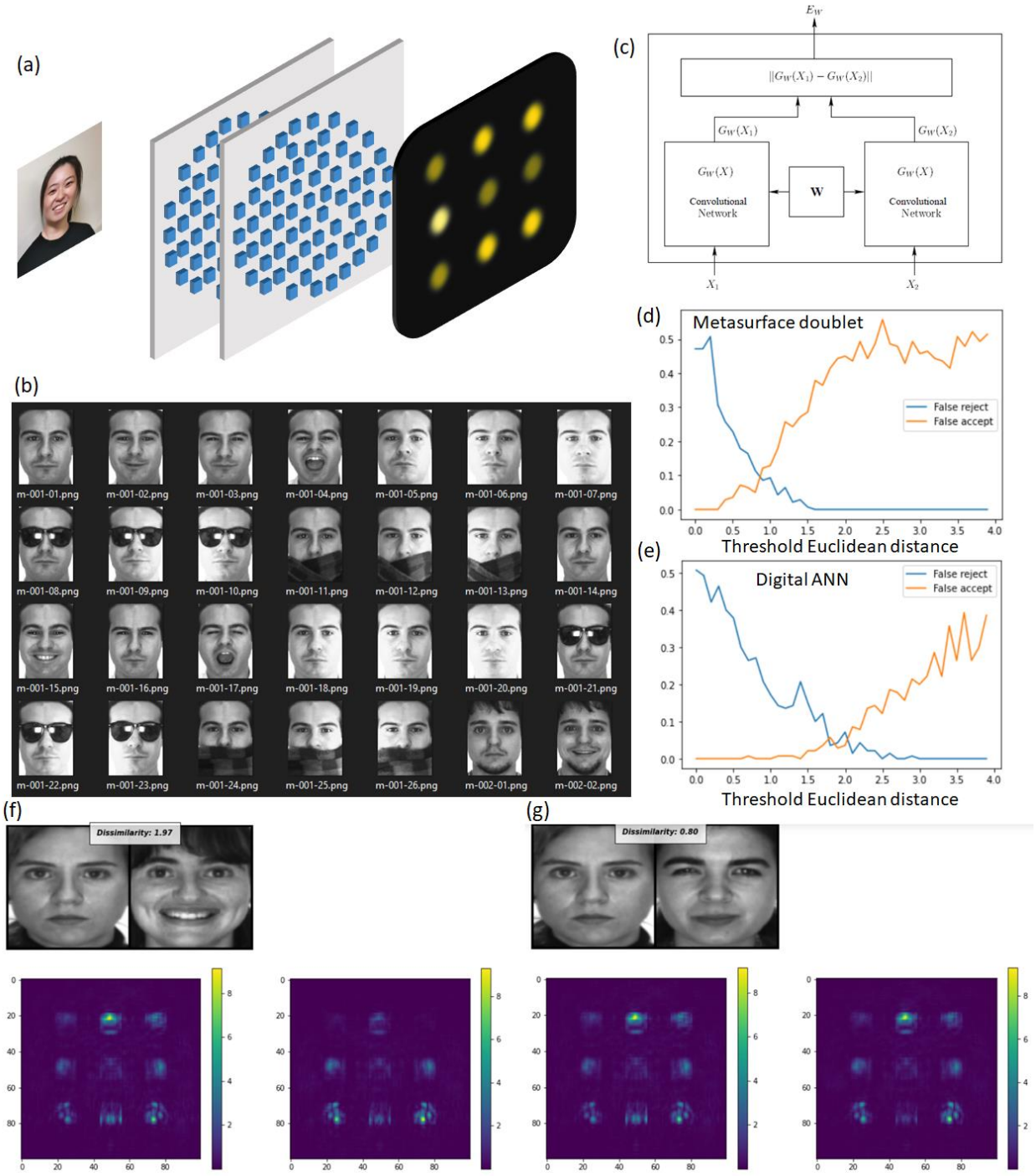

Figure 3.5 **Metasurface doublet for human face verification.** (a) Schematic illustrating the working mechanism of the metasurface doublet. An input gray-scale human face image is mapped into an array (intensity pattern over an array of predefined zones on the camera plane) and the similarity of two images is evaluated by comparing the Euclidean distance between their corresponding arrays to determine if they represent the same person or not. (b) Example human face photos used in designing and testing the doublet smart glass. Only photos without facial cover (e.g., images #1-7, and #14-20) were used. (c) Siamese neural network used for our human face verification task. (d) Results showing that when the threshold Euclidean distance was taken to be ~0.8, both the false rejection rate and false acceptance rate of our metasurface face verification system are reduced to ~10%. (e) Results showing that in a control digital ANN with 3 convolutions and 2 fully connected layers, when the threshold Euclidean distance was taken to be ~1.8, both the false rejection rate and false acceptance rate are reduced to ~5%. (f) An example result showing that two photos were determined to not represent the same person because their Euclidean distance determined by the metasurface doublet is above the threshold. (g) An example result showing that two photos were determined to represent the same person because their Euclidean distance determined by the metasurface doublet is below the threshold.

3.2 Resonant Recurrent Neural network

3.2.1 Acoustic resonators to recognize vowel sound

Figure 3. (b) and **(c)** show the results of the cross-entropy loss and the prediction accuracy, respectively. The trained resonance system can achieve an accuracy of 81.7% for the training dataset and 83.3% accuracy for the test dataset over 50 training epochs. The confusion matrix for the test data indicates that the resonance system can indeed perform vowel recognition [inset of **Figure 3. (c)**]. The time-integrated power at each waveguide demonstrates that the optimized

resonant architecture can route most of the energy of the object vowel class to the correct channel [**Figure 3. (d)**].

The classification of vowels requires memory in the temporal domain. Resonance provides an explicit form of analog memory, making it possible to directly interpret long and short-term memory effects. To demonstrate that, we consider a challenging case in RNNs: long short-term memory (LSTM). LSTM is designed to solve the problem of learning long-term temporal dependencies. The gradient of the loss function decays exponentially with time, causing the issue of vanishing gradient. LSTM is explicitly designed to include both short-term and long-term memory. A common LSTM architecture is shown in Fig. 11(a). It incorporates cell states and gates

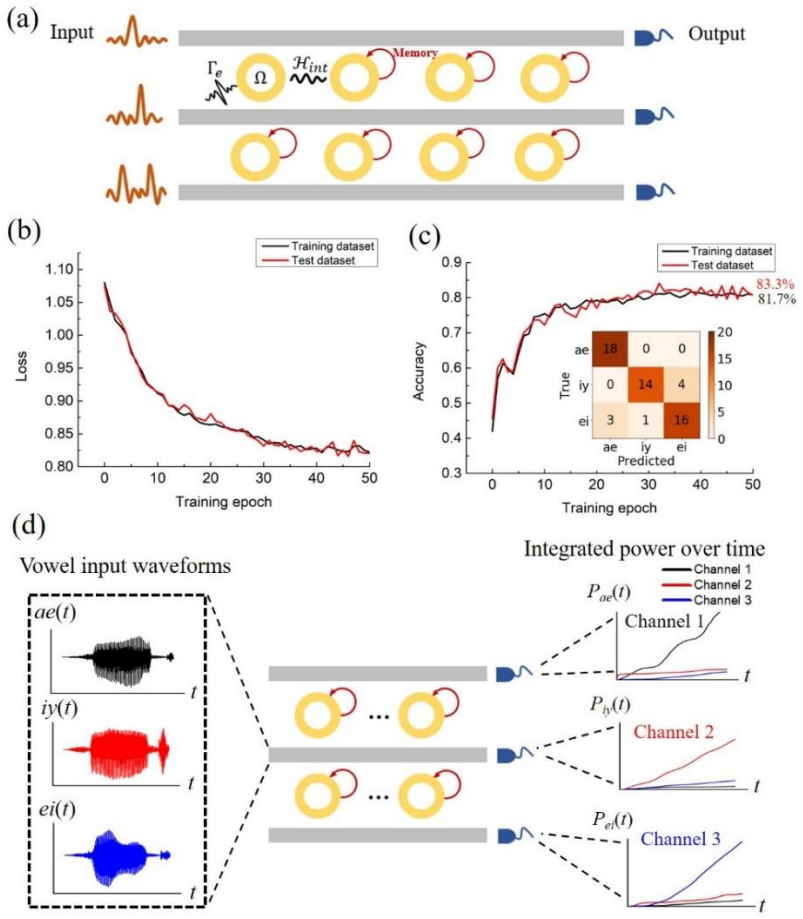

Figure 3.6 **Acoustic resonators to recognize vowel sound in the wave domain.** (a) Schematic of the vowel-recognition resonant recurrent network with three waveguides and two rows of resonators. (b) Cross-entropy loss and (c) prediction accuracy over 50 training epochs. Inset is confusion matrix for the test dataset. The correct predictions are in the diagonal of the table and prediction errors are located outside the diagonal. (d) Raw audio waveforms of three spoken vowel classes and time-integrated power at each output channel.

such that events from the remote past can have current impacts. It is difficult to construct and interpret LSTM in existing wave computing systems; however, in the resonance system, we can include resonators with different lifetimes to accomplish short-term and long-term memory.

3.2.2 The task of long-term and short-term memory

We now address a more challenging task by making the previous task of vowel classification more practical. Previously, the input signal consisted of a pure vowel without any quiet lapse of time preceding or following the vowel. This would be an unusual situation. In practice, we often take a temporal window of recorded sound and ask if there is a vowel in this window. The temporal window is often much longer than the duration of the vowel. Now we need relatively short-term memory to recognize the inner structure of the vowel sound, and at the same time, we also need long-term memory because the duration of the temporal window can be much longer than the length of the vowel sound.

The specific case study considered here uses a temporal window of 600-time steps. A vowel sound only consists of 100-time steps. We position the vowel at different temporal locations as shown in **Figure 3. (c-e).** Here we use voices from three classes: *ae*(*t*), *iy*(*t*), and *ei*(*t*). In the first training case, we place a vowel sound at the end of the temporal window [**Figure 3. (c)**]. In the second case, we place it in the middle, starting at time step number 300 [**Figure 3. (d)**]. And lastly, we place it towards the beginning of the temporal window [**Figure 3. (e)**]. The neural network makes

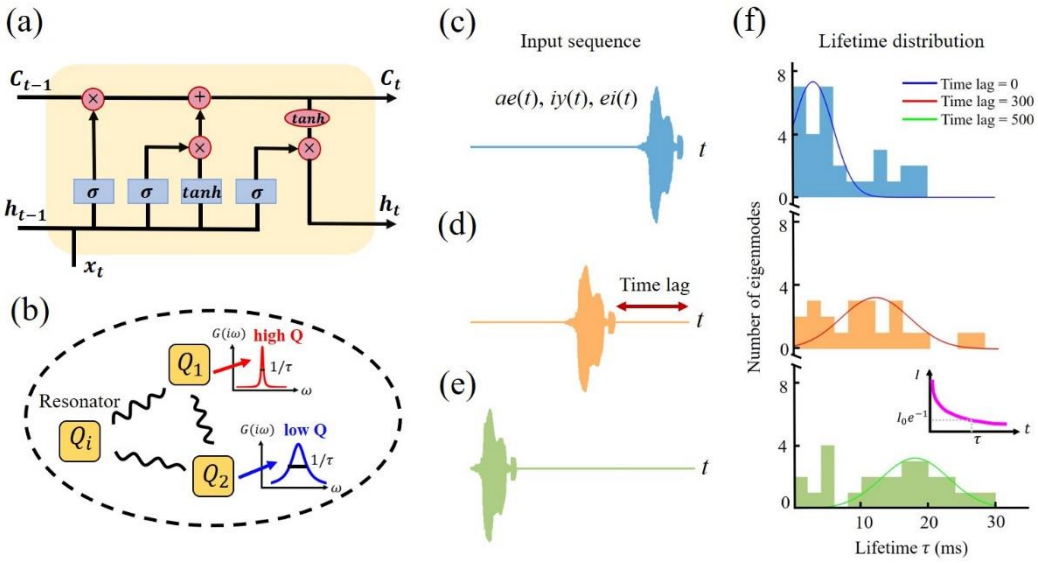

Figure 3.7 **Task of long-term and short-term memory.** (a) Diagram of LSTM architecture in a digital RNN. (b) Diagram of coupled resonators with different quality factors Q and lifetimes τ. (c-e) Schematics showing that the timing of vowel sound varies for three different training sets. The vowel signal contains three classes: *ae*(*t*), *iy*(*t*) and *ei*(*t*). (f) The lifetime distribution of resonators after training for three cases. Long-term and short-term memory arise spontaneously in response to different memory requirement in three different training settings.

its decision of the vowel class at the end of temporal window. For the first case, the vowel sound is located at the end of the temporal window. There is no time lag between the vowel sound and the output so that the network requires mostly short-term memory. For the last case, however, there is a long-time lag, and therefore long-term memory is needed. This setup allows us to see how short and long-term memories spontaneously arise from training.

The three trained computing systems all function well with accuracies above 70%. It is interesting to examine the lifetimes of resonators in the three computing systems. Because of the coupling between resonators, it is more informative to look at the lifetimes of eigenmodes after modal hybridization. Eigenvalues of the matrix $\mathcal{H} = i\Omega - \Gamma_e - \mathcal{H}_{int}$ are calculated, where the imaginary and real parts are the resonant frequency ω and decay rate $1/\tau$ of each eigenmode, respectively. The lifetime τ of each eigenmode can then be calculated. The histogram of eigenmodes shows distinct lifetime distributions for the three cases [**Figure 3. (f)**]. For the first case, the computing system consists of resonators mostly with short lifetimes. For the second case, the histogram shows a middle-range lifetime distribution. For the third case with the longest time lag, the resonance system evolves to possess the capability of long-term memory by including a significant portion of resonators with long lifetimes; in fact, the computing system develops both short-term and long-term memory simultaneously, in a similar way as the memory cells in LSTM. Thus, we observe that the distribution of lifetime semi-quantitatively reflects the length of the memory required in each computing situation.

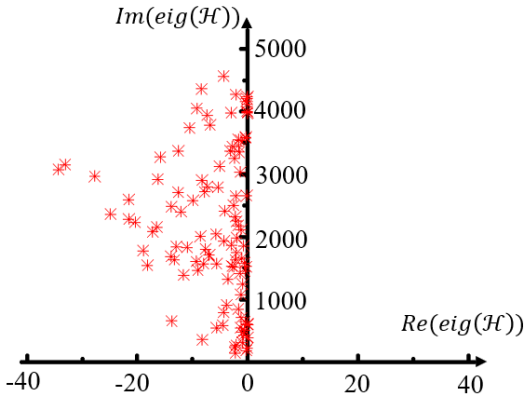

Figure 3.8 **Real and imaginary parts of the eigenvalues of matrix $\mathcal{H}$**. The training case we studied here is the same as the one in Figure 3.

We note that the resonance systems have many intrinsic advantages over digital RNNs. These include advantages associated with analog wave computing, such as fast speed, ultra-low energy consumption, and intrinsic parallelism. Moreover, the proposed system can process continuous temporal signals whereas digital RNNs have the risk of insufficient temporal resolution. Lastly, an important but subtle advantage is that it is much easier to train a resonance system than to train a digital RNN. Training a digital RNN has an inherent issue of instability known as the problem of

exploding gradient, which is due to the great depth of RNNs. One has to take great care to address exploding gradient problems by using techniques such as gradient clipping and weight regularization. It has been shown that in a physical system, recurrent computing is fundamentally stable. We can quantify the advantage of training stability by treating the computing device as a dynamic system described by $\dot{a}(t) = \mathcal{H} \cdot a(t)$. This dynamic system is stable if all the eigenvalues of the matrix $\mathcal{H}$ have negative real parts. As shown in **Figure 3.**, all eigenvalues of the trained matrix $\mathcal{H}$ are in the left half the space; that is, the training of this computing system is fundamentally stable.

4 CONCLUSION

4.1 Smart Glass

We have successfully demonstrated a novel approach to implement optical neural networks capable of recognizing simple objects with high accuracy. Optimization of the ONN configurations, for example, the size, shape, and arrangement of detection domains, can lead to higher accuracy in object recognition. By exploiting amplitude-phase modulation in metasurfaces, compound metasurface (e.g., doublets, triplets), and wavelength multiplexing, the smart glass can have more degrees of freedom in the control of network parameters to perform more complex recognition tasks. Our metasurface ONNs and ONNs based on integrated photonics have only implemented the linear transformation operation, not the nonlinear activation operation, and in this sense, they are not exact replicants of their biological counterparts. This fact limits the range of tasks they can perform and the accuracy they can achieve. Future work will realize nonlinear activation by introducing nonlinear materials (e.g., semiconductors with saturable absorption) into metasurfaces. To enhance optical nonlinearity, the inputs can take the form of short pulses and resonant structures can be incorporated into the ONN.

Advanced sensors will be ubiquitous in future applications. These sensors are often deployed in areas or scenarios that lack infrastructure support. They require minimal service, resilience to influence, high energy efficiency, and information security. These requirements present a daunting challenge for existing technology. Today, collecting data in the physical domain involves a complex technology stack: a device that measures physical input, analog-to-digital conversion, and digital processors. The system ends up being vulnerable to influence, power-hungry, and reacting slowly due to the latency between modules. An ONN could fundamentally reshape the future of data collection and analysis by developing a new paradigm of “edge” perception devices. It computes directly upon the physical domain, effectively condensing measurement, A/D conversion, and computing in a single passive device. It uses no power or orders of magnitude less power, provides physics-guaranteed security, and comes in an extremely robust and ultra-compact form factor. Importantly, it can protect the privacy of the subject of interest because there is no representation of the subject in the digital domain. For example, we can envision, a standalone biometric lock composed of a customized smart glass with a few photodetectors and a coherent light source to recognize human fingerprints or facial profiles.

4.2 Resonant Recurrent Neural network

In conclusion, we have shown that resonance can be used to construct stable and scalable recurrent neural networks. Resonance provides an explicit form of memory. Short-term or long-term memories can be directly constructed and interpreted by the lifetime of resonators. This extends the analog computing capability into the complex neural network architecture such as LSTM. While we use acoustic as an example, the strategy can be broadly applied to other physical domains. Photonic analog computing has been used in several applications such as reservoir computing [12], photonic Ising machines [13], self-learning Machines [14], image edge detection [15], and analog signal processing [16].

REFERENCES

[1] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nat. Photonics*, vol. 11, no. 7, pp. 441–446, Jul. 2017, doi: 10.1038/nphoton.2017.93.

[2] X. Lin *et al.*, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, Sep. 2018, doi: 10.1126/science.aat8084.

[3] E. Khoram, A. Chen, D. Liu, Q. Wang, M. Yuan, and Z. Yu, "Optimization of Nonlinear Nanophotonic Media for Artificial Neural Inference," in *Conference on Lasers and Electro-Optics (2019), paper JM3M.4*, May 2019, p. JM3M.4. doi: 10.1364/CLEO_AT.2019.JM3M.4.

[4] N. Yu *et al.*, "Light Propagation with Phase Discontinuities: Generalized Laws of Reflection and Refraction," *Science*, vol. 334, no. 6054, pp. 333–337, Oct. 2011, doi: 10.1126/science.1210713.

[5] R. J. Lin *et al.*, "Achromatic metalens array for full-colour light-field imaging," *Nat. Nanotechnol.*, vol. 14, no. 3, pp. 227–231, Mar. 2019, doi: 10.1038/s41565-018-0347-0.

[6] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, no. 5, pp. 602–610, Jul. 2005, doi: 10.1016/j.neunet.2005.06.042.

[7] G. K. Anumanchipalli, J. Chartier, and E. F. Chang, "Speech synthesis from neural decoding of spoken sentences," *Nature*, vol. 568, no. 7753, pp. 493–498, Apr. 2019, doi: 10.1038/s41586-019-1119-1.

[8] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to Sequence Learning with Neural Networks," in *Advances in Neural Information Processing Systems*, 2014, vol. 27. Accessed: Dec. 05, 2021. [Online]. Available: https://papers.nips.cc/paper/2014/hash/a14ac55a4f27472c5d894ec1c3c743d2-Abstract.html

[9] J. L. Elman, "Finding structure in time," *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, Apr. 1990, doi: 10.1016/0364-0213(90)90002-E.

[10] T. W. Hughes, I. A. D. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," *Sci. Adv.*, vol. 5, no. 12, p. eaay6946, doi: 10.1126/sciadv.aay6946.

[11] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[12] G. Marcucci, D. Pierangeli, and C. Conti, "Theory of neuromorphic computing by waves," in *OSA Nonlinear Optics 2021 (2021), paper NTh1A.7*, Aug. 2021, p. NTh1A.7. Accessed: Dec. 05, 2021. [Online]. Available: https://www.osapublishing.org/abstract.cfm?uri=NLO-2021-NTh1A.7

[13] C. Roques-Carmes *et al.*, "Heuristic recurrent algorithms for photonic Ising machines," *Nat. Commun.*, vol. 11, no. 1, p. 249, Jan. 2020, doi: 10.1038/s41467-019-14096-z.

[14] V. Lopez-Pastor and F. Marquardt, "Self-learning Machines based on Hamiltonian Echo Backpropagation," *ArXiv210304992 Nlin Physicsphysics*, Mar. 2021, Accessed: Dec. 05, 2021. [Online]. Available: http://arxiv.org/abs/2103.04992

[15] T. Zhu *et al.*, "Generalized Spatial Differentiation from the Spin Hall Effect of Light and Its Application in Image Processing of Edge Detection," *Phys. Rev. Appl.*, vol. 11, no. 3, p. 034043, Mar. 2019, doi: 10.1103/PhysRevApplied.11.034043.

[16] F. Zangeneh-Nejad and R. Fleury, "Topological analog signal processing," *Nat. Commun.*, vol. 10, no. 1, p. 2058, May 2019, doi: 10.1038/s41467-019-10086-3.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

AFRL	Air Force Research Laboratory
ANNs	Artificial neural networks
ICP	inductively coupled plasma
LSTM	Long short-term memory
ONNs	optical neural networks
RCWA	Rigorous Coupled Wave Analysis
RNNs	recurrent neural networks
RXAS	Soft Matter Materials Branch, Materials and Manufacturing Directorate
SEM	Scanning electron microscope