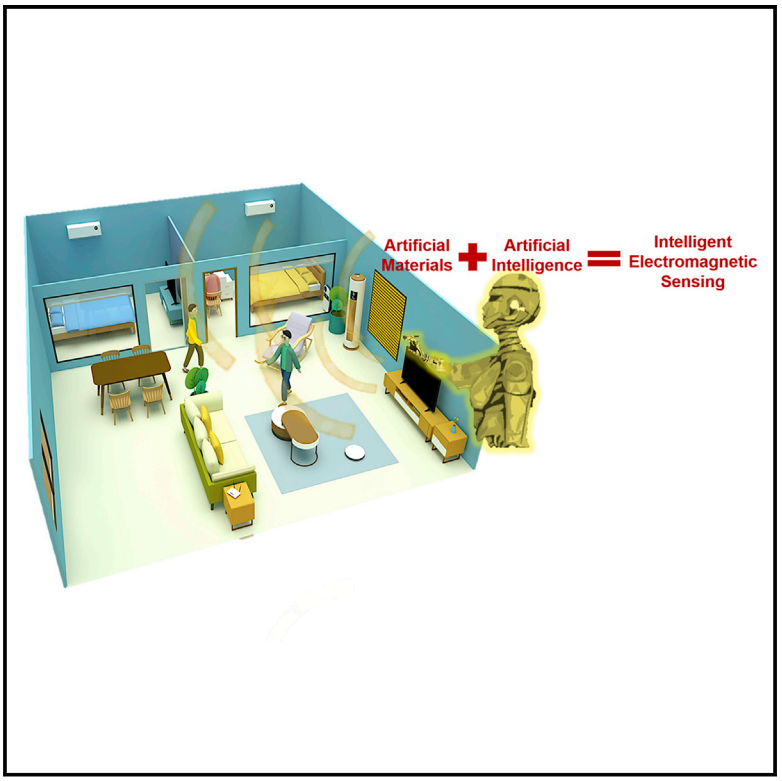
Article

Patterns

# Intelligent Electromagnetic Sensing with Learnable Data Acquisition and Processing

## Graphical Abstract




## Highlights

- First demonstration of "learned sensing" in electromagnetics
- Low-latency imaging and gesture recognition with learned illumination patterns
- Data acquisition and processing integrated into unique machine-learning pipeline
- Selection of task-relevant information already during measurements

## Authors

Hao-Yang Li, Han-Ting Zhao, Meng-Lin Wei, ..., Tie Jun Cui, Philipp del Hougne, Lianlin Li

## Correspondence

tjcui@seu.edu.cn (T.J.C.), philipp.delhougne@gmail.com (P.d.H.), lianlin.li@pku.edu.cn (L.L.)

## In Brief

"Smart" devices must "see" and "recognize" objects and gestures in their surroundings as quickly as possible. We consider a contactless sensor that illuminates its surroundings with microwave illumination shaped by a programmable metasurface. By integrating the measurement process directly into the machine-learning pipeline that processes the data, we learn optimal illumination patterns that efficiently extract task-relevant information. Our experimental demonstration of low-latency intelligent electromagnetic sensing will influence human-computer interaction, health care, automotive radar, and security screening.

Li et al., 2020, Patterns *1*, 100006
April 10, 2020 
https://doi.org/10.1016/j.patter.2020.100006

CellPress

Patterns

CellPress
OPEN ACCESS

Article

# Intelligent Electromagnetic Sensing with Learnable Data Acquisition and Processing

Hao-Yang Li,[1] Han-Ting Zhao,[1] Meng-Lin Wei,[1] Heng-Xin Ruan,[1] Ya Shuang,[1] Tie Jun Cui,[2,*] Philipp del Hougne,[3,*] and Lianlin Li[1,4,*]

[1]State Key Laboratory of Advanced Optical Communication Systems and Networks, Department of Electronics, Peking University, Beijing 100871, China

[2]State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China

[3]Institut de Physique de Nice, CNRS UMR 7010, Université Côte d'Azur, Nice 06108, France

[4]Lead Contact

*Correspondence: tjcui@seu.edu.cn (T.J.C.), philipp.delhougne@gmail.com (P.d.H.), lianlin.li@pku.edu.cn (L.L.)

https://doi.org/10.1016/j.patter.2020.100006

**THE BIGGER PICTURE** Many futuristic "intelligent" concepts that will affect our society, from ambient-assisted health care via autonomous vehicles to touchless human-computer interaction, necessitate sensors that can monitor a device's surroundings fast and without extensive computational effort. To date, sensors indiscriminately acquire *all* information and only select relevant details during data processing, thereby wasting time, energy, and computational resources. We demonstrate intelligent electromagnetic sensing that uses learned illumination patterns to already select *relevant* details during the measurement process. Our experiments use a home-made programmable metasurface to generate the learned microwave patterns that enable a remarkable reduction in the number of necessary measurements. Our demonstration addresses a widespread need for high-quality contactless electromagnetic sensing under strict time, energy, and computation constraints.


**Development/Pre-production:** Data science output has been rolled out/validated across multiple domains/problems

## SUMMARY

Electromagnetic (EM) sensing is a widespread contactless examination technique with applications in areas such as health care and the internet of things. Most conventional sensing systems lack intelligence, which not only results in expensive hardware and complicated computational algorithms but also poses important challenges for real-time *in situ* sensing. To address this shortcoming, we propose the concept of intelligent sensing by designing a programmable metasurface for data-driven learnable data acquisition and integrating it into a data-driven learnable data-processing pipeline. Thereby, a measurement strategy can be learned jointly with a matching data post-processing scheme, optimally tailored to the specific sensing hardware, task, and scene, allowing us to perform high-quality imaging and high-accuracy recognition with a remarkably reduced number of measurements. We report the first experimental demonstration of "learned sensing" applied to microwave imaging and gesture recognition. Our results pave the way for learned EM sensing with low latency and computational burden.

## INTRODUCTION

Electromagnetic (EM) sensing is a widely used contactless examination technique because it relies on harmless nonionizing radiation that can penetrate optically opaque materials. Consequently, EM sensing has emerged as promising tool in various applications ranging from health care[1–4] via security screening[5–7] up to planetary explorations.[8–10] However, to fully exploit the application potential of EM sensing, cost efficiency and speed remain two major challenges. Traditional EM hardware relies on mechanically or electronically scanned beams, suffering from slow acquisition or high cost, respectively. Recently, novel hardware solutions to shift the cost from the hardware to the software level were proposed: rather than acquiring the information in the spatial domain, the information can be multiplexed across independent frequencies or configurations of a complex system such as a chaotic cavity or a metasurface aperture.[11–14] The computationally intensive challenge is then to recover the spatial information from the multiplexed measurements. Given the typical scene sparsity, compressive

sensing schemes are usually employed to recover the information from an underdetermined set of measurements.[15] Especially when analytical models are lacking, machine-learning (ML) techniques have recently been proposed to solve the electromagnetic inverse scattering problem.[16–18]

However, these EM sensing schemes based on generic (random) scene illuminations still lack "intelligence": they indiscriminately acquire all information, ignoring available knowledge about scene, sensing task, and hardware constraints. Yet using this available a priori knowledge is critical to limit data acquisition to *relevant* information—the crucial conceptual improvement necessary to reduce latency and computational burden. A first step to add intelligence consisted in adapting the scene illuminations to the knowledge of what scene is expected via a principal component analysis (PCA) of the expected scene.[19,20] Albeit yielding significant performance improvements, this technique still considers acquisition and processing separately, and hence fails to use task-specific measurement settings that highlight salient features for the processing layer.

To fully reap the benefits of ML techniques in EM sensing, acquisition and processing must be jointly learned in a unique pipeline. Recently, inspired by pioneering work in optical microscopy,[21] the idea of learned EM sensing with programmable metasurface hardware was introduced.[22] Thereby, a model of the programmable measurement process is directly integrated into the ML pipeline used to process the data, enabling the joint learning of optimal measurement and processing settings for the given hardware, task, and expected scene. We note that there is a number of related works in optics[23–27] and a similar concept has recently also been studied in the context of ultrasonic imaging.[28] It is important to realize that most of the time in EM sensing, the specific task can be executed without first reconstructing a visual image of the scene. While current EM sensing systems are designed to first generate high-resolution images as a checkpoint for subsequent ML-aided recognition tasks, the resulting acquisition of irrelevant information and the huge flux of data to be processed cause significant inefficiencies. Hence, for tasks such as object recognition it is most efficient to skip the intermediate imaging step and directly process the raw data, as done in previous studies.[20,22]

Here, we report the first experimental demonstration of learned EM sensing. We consider the tasks of (1) imaging and (2) gesture recognition, two examples with immediate real-life application in security screening and human-machine interaction. Our intelligent hardware layer relies on a programmable metasurface[29–31] to shape the waves illuminating the scene. Unlike del Hougne et al.,[22] we do not use an analytical description of the measurement process but train a three-port deep artificial neural network (ANN), called the measurement ANN (m-ANN), to capture the link between scene, metasurface coding pattern, and microwave measurement. Being sharply different from the study by Li et al.,[32] where only the data processing is treated with the ANN, here we jointly optimize the control coding patterns of the metasurface together with the weights of another ANN, called the reconstruction ANN (r-ANN), which is used to extract the desired information from the raw measurements. To this end, we interpret measurement and reconstruction as, respectively, encoding and decoding the relevant scene information (an image or a gesture classification) in/from a measurement space. Thus, we can employ a variational autoencoder (VAE) framework[33–35] to jointly learn optimal measurement and processing settings with a standard supervised-learning procedure. This strategy drastically reduces the number of necessary measurements, helping us to remarkably improve many critical metrics such as speed, processing burden, and energy consumption.

## RESULTS

### Operation Principle

Learned sensing requires the integration of a model of the reconfigurable measurement process into the pipeline used for data processing, in order to jointly optimize the learnable physical and digital weights, as illustrated in Figure 1. Given the difficulty of accurately modeling the measurement process in an analytical manner, for instance, due to inaccuracies in the metasurface fabrication or multipath effects in the indoor environment, here we propose the use of a data-driven model of the measurement process instead. We introduce a three-port-deep ANN, the measurement ANN (m-ANN), that links two inputs (the scene $\boldsymbol{x}$ and the metasurface configuration $\mathcal{C}$) to one output (the raw microwave measurements $\boldsymbol{y}$). First, we learn the trainable weights $\boldsymbol{\Theta}$ of the m-ANN from an $(\boldsymbol{x}, \mathcal{C}, \boldsymbol{y})$-triple labeled training dataset with a standard supervised-learning procedure, as detailed in Experimental Procedures and Note S4.

Once the weights $\boldsymbol{\Theta}$ of m-ANN are fixed, we can proceed with the integration of m-ANN into the unique sensing pipeline. Following Vedula et al.,[28] we view the entire sensing process (data acquisition and processing) as a user-controlled end-to-end process. Given a scene $\boldsymbol{x}$ (an image or a gesture class), a set of complex-valued measurements $\boldsymbol{y}$ is generated by sampling from the $\mathcal{C}$-controllable conditional distribution $\boldsymbol{y} \sim q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Theta})$. In other words, the latent scene variable of interest, $\boldsymbol{x}$, is *encoded* in a measurement space $\boldsymbol{y}$ via a distribution controlled by the metasurface configuration $\mathcal{C}$. The goal of the processing is then to find an estimator that retrieves the relevant scene information $\boldsymbol{x}$ from the measurements $\boldsymbol{y}$. This estimator inverts the action of the measurement process—in other words, it *decodes* the information of interest to return from the measurement space to the latent variable space. Using the VAE[33–35] framework, the digital decoder can be modeled as sampling the measurement space with a conditional distribution $\boldsymbol{x} \sim p_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y})$, controlled by its digital weights $\boldsymbol{\Phi}$, to generate estimates of the latent variable of interest. The decoding is implemented with a deep ANN, called r-ANN, whose trainable weights can hence be identified as $\boldsymbol{\Phi}$.

To jointly learn optimal analog and digital weights, i.e., the metasurface control coding pattern $\mathcal{C}$ and the r-ANN weights $\boldsymbol{\Phi}$, respectively, we minimize the following objective function:[35]

$$\mathcal{L}(\mathcal{C}, \boldsymbol{\Phi}) = -\mathbb{E}_{q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})}[\log p_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y})] + \mathrm{KL}(q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Theta})|p(\boldsymbol{y})). \quad \text{(Equation 1)}$$

The first term in Equation 1, $-\mathbb{E}_{q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})}[\log p_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y})]$, can be interpreted as the "reconstruction error" of our VAE: it is the log likelihood of the true latent data given the inferred latent data. The second term, $\mathrm{KL}(q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\Theta})|p(\boldsymbol{y}))$, acts as a regularizer and encourages the distribution of the decoder to be close to a

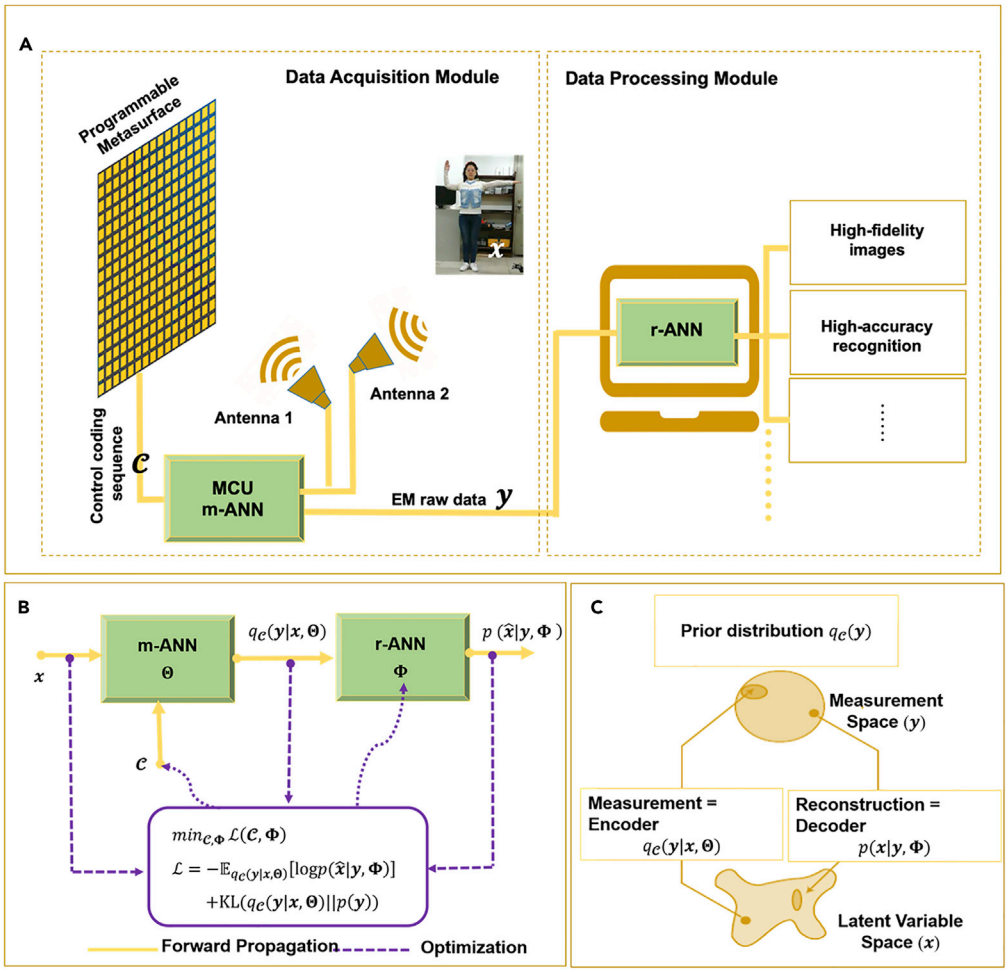



**Figure 1. Setup and Working Principle of the Learned EM Sensing System**

(A) The intelligent sensing system consists of two data-driven learnable modules: the m-ANN-driven data acquisition module and the r-ANN-driven data processing module. m-ANN models the measurement process involving a pair of horn antennas and a coding metasurface programmed with a micro-control unit. Antenna 1 emits plane microwave fronts, which are shaped by the programmable metasurface and then incident upon the scene. The waves scattered by the subject of interest are received by Antenna 2. The received raw microwave data are directly processed by the r-ANN, producing the desired imaging or recognition results.

(B) The m-ANN-based model of data acquisition as a trainable physical network is fully integrated with the r-ANN-driven data-processing pipeline into a unique sensing chain. The m-ANN has three ports: the scene of interest $\boldsymbol{x}$ and the metasurface coding pattern $\mathcal{C}$ that shapes the scene illumination are its inputs, and the raw EM data $\boldsymbol{y}$ is its output. First, the weights $\boldsymbol{\Theta}$ of m-ANN are fixed using a data-driven supervised-learning protocol. Second, optimal values of the physical and digital parameters of the sensing chain, $\mathcal{C}$ and $\boldsymbol{\Phi}$, respectively, are jointly learned using a standard supervised-learning procedure (error back-propagation) in a VAE framework. These optimal settings ensure that already on the hardware level only task-relevant information is captured and that a matching processing layer extracts the relevant information from the measurements.

(C) Our interpretation of the entire sensing process in the VAE framework: the latent variable space, $\boldsymbol{x}$, is encoded by the analog measurements in a measurement space, $\boldsymbol{y}$; the digital reconstruction decodes the measurements to return to the latent variable space.

chosen prior distribution $p(\boldsymbol{y})$. In our work, both analog encoder $q_{\mathcal{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})$ and digital decoder $p_{\boldsymbol{\Phi}}(\boldsymbol{x}|\boldsymbol{y})$ are treated with deep ANNs, namely m-ANN and r-ANN, as detailed in Experimental Procedures and Notes S1 and S4.

To determine the optimal settings of $\mathcal{C}$ and $\boldsymbol{\Phi}$, we apply a so-called alternatively iterative approach. Starting with some initializations of $\mathcal{C}$ and $\boldsymbol{\Phi}$, we calculate $\mathcal{C}$ (resp. $\boldsymbol{\Phi}$) for $\boldsymbol{\Phi}$ (resp. $\mathcal{C}$) updated in the last iteration step, followed by calculating $\boldsymbol{\Phi}$ (resp. $\mathcal{C}$) based on the obtained $\mathcal{C}$ (resp. $\boldsymbol{\Phi}$). This procedure is repeated until a stopping criterion is fulfilled. Such an optimization can be implemented using error-backpropagation[36] routines for continuous variables such as $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$. However, we have a binary constraint (0 or 1) on the entries of $\mathcal{C}$, resulting in an NP-hard problem. While del Hougne et al.[22] addressed this binary constraint with a temperature parameter trick,[23] in this work we use a randomized simultaneous perturbation stochastic approximation (r-SPSA) algorithm. The latter was originally developed for treating the problem of optimal well placement and control in the area of petroleum engineering.[37] More details about Equation 1 and the implementation of the optimization algorithm are provided in Figures S2 and S3 and Note S2.

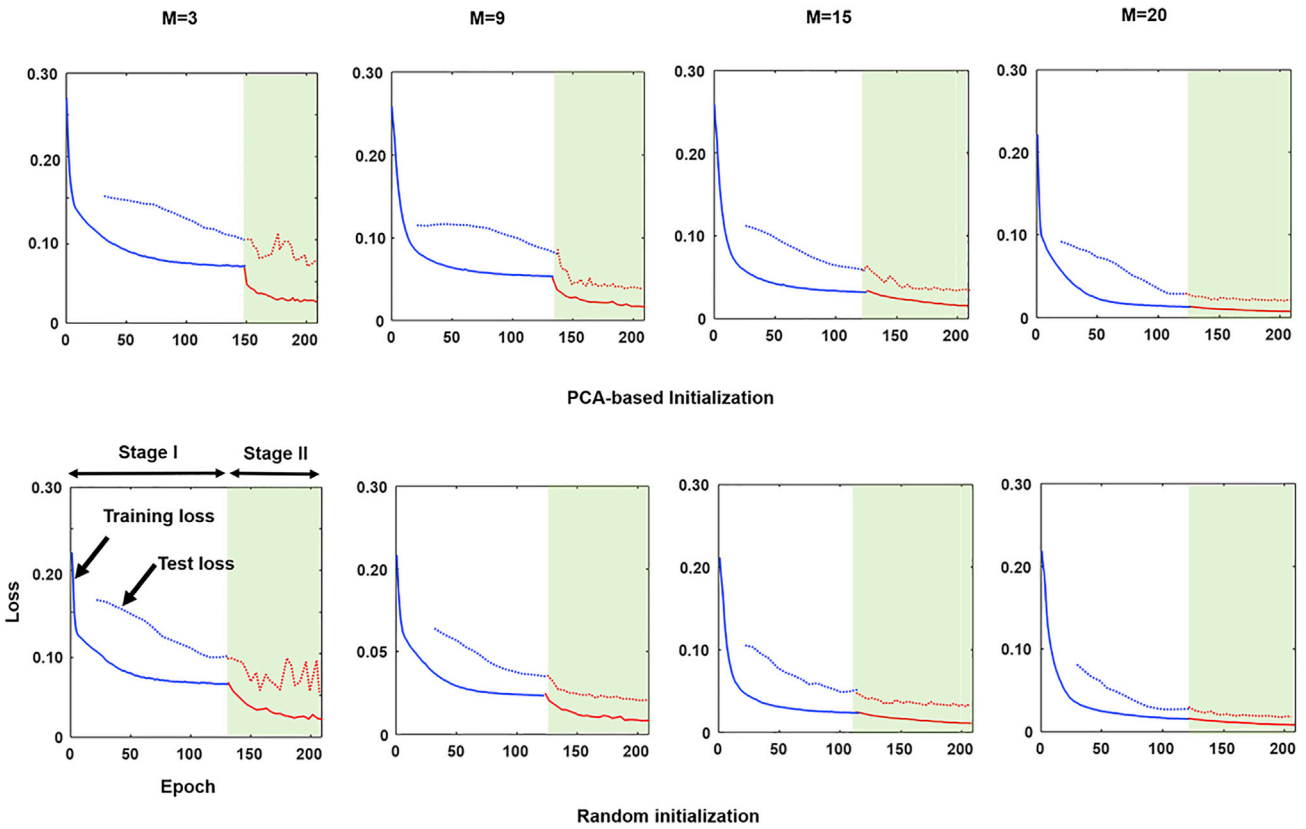



**Figure 2. Training Dynamics for Learned EM Sensing Applied to an Imaging Task**

The dependence of the training and test loss functions on the progress of iterative epochs is shown for different numbers of coding patterns M, i.e., M = 3, 9, 15, and 20. The continuous lines indicate the training loss and the dashed lines indicate the test loss. The control coding patterns of the metasurface are initialized randomly (top) or PCA-based (bottom). During stage I, only the digital decoder weights $\mathbf{\Phi}$ are optimized. Then, during stage II, both the physical weights $\mathcal{C}$ and the digital weights $\mathbf{\Phi}$ are jointly optimized. The presented results show a remarkable improvement of the image quality achieved by using the joint optimization of $\mathcal{C}$ and $\mathbf{\Phi}$ during stage II, compared with that based on solely optimizing $\mathbf{\Phi}$ (i.e., the end of stage I). The effect is especially striking when the number of measurements is very limited.

### *In Situ* Imaging of the Human Body

First, we apply our learned EM sensing system to the task of *in situ* high-resolution imaging of a human body in our laboratory environment. As outlined previously, we integrate m-ANN for data acquisition and r-ANN for smart data processing into a unique data-driven learnable sensing chain. To this end, we need to jointly optimize the coding patterns $\mathcal{C}$ of the m-ANN together with the weights $\mathbf{\Phi}$ of r-ANN for the specific task of human body imaging. The integrated ANN, containing a multitude of nonlinear ML layers, can be trained with a standard supervised-learning procedure in TensorFlow. Following Vedula et al.,[28] to illustrate the significant improvement of the proposed learned sensing strategy on the image quality over conventional learning-based sensing methods, we consider a two-stage training procedure. During the first stage, the coding patterns of the m-ANN and the digital weights of the r-ANN are optimized separately, as in Li et al.[20] The coding patterns of m-ANN are assigned following the two most common state-of-the-art approaches that correspond to using random or PCA-based scene illuminations. During the second stage, m-ANN and r-ANN are jointly trained to achieve the overall optimal sensing performance. In this two-stage training, the benefit reaped by the proposed sensing strategy over the conventional methods can be clearly demonstrated. In our study we use several people, called training persons in short, to train our intelligent microwave sensing system, and use a different person, called test persons, to test it. In addition, we use 1,000 random codes and 1,000 PCA-based codes (200 standard PCA-based codes and their 800 perturbations) as raining samples for training $\mathbf{\Theta}$. The details of the training persons are provided by Li et al.[32] The ground truth is defined using binarized optical images of the scene.

In general, depending on the difficulty of the sensing task and the signal-to-noise ratio,[38] a measurement with a single coding pattern cannot be expected to obtain sufficient relevant information. Figure 2 displays the cross-validation errors over the course of the training iterations for different numbers M of coding patterns of the metasurface (3, 9, 15, and 20). The two stages of the aforementioned training protocol can be clearly distinguished. To assess the image quality quantitatively, the so-called structure similarity index metric (SSIM) is considered, which is calculated using the MATLAB library function, i.e., ssim. Figure S4 displays SSIM histograms for different values of M and different sensing methods. Since the trainable physical ($\mathcal{C}$) and digital ($\mathbf{\Phi}$) parameters are initialized randomly before

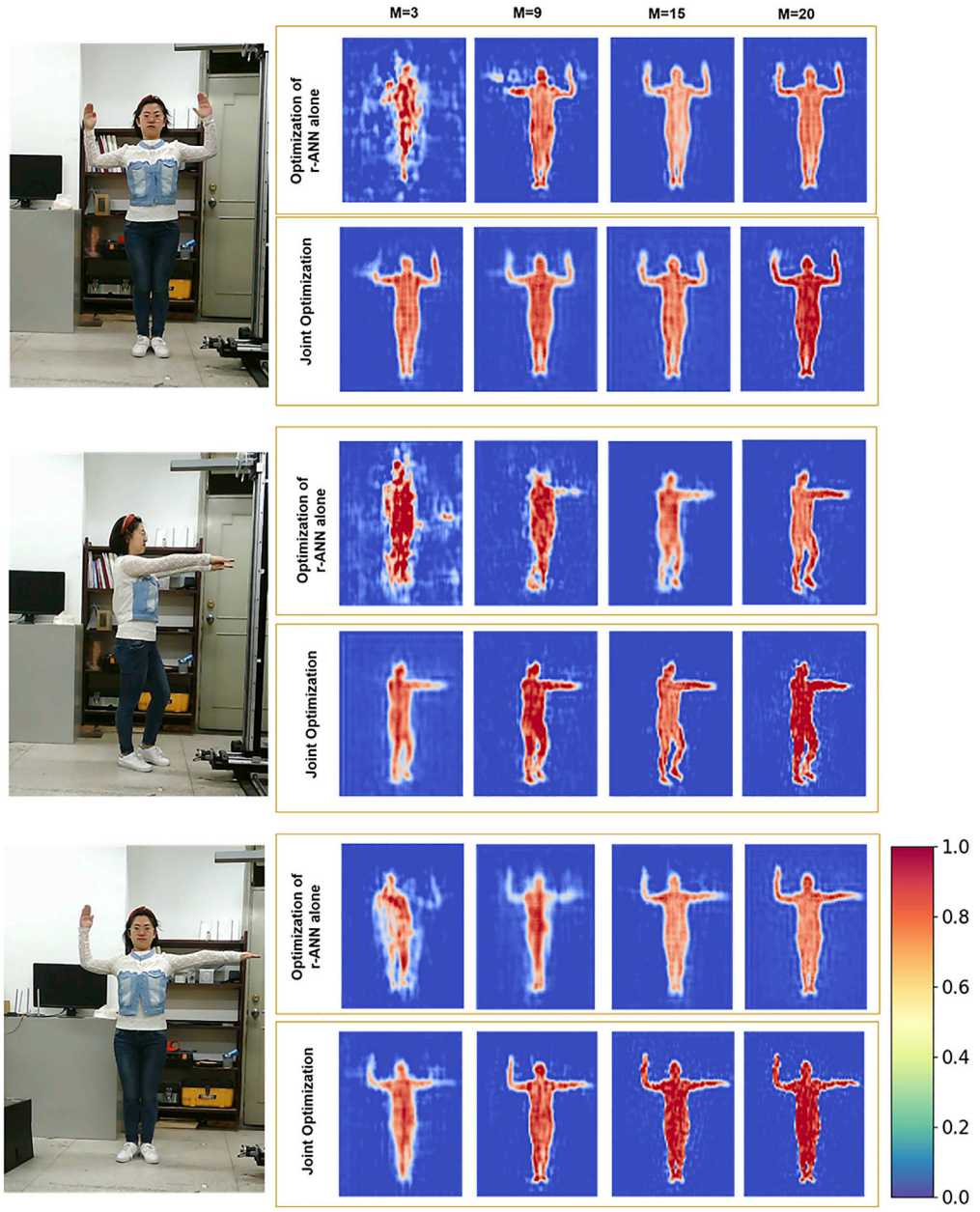



**Figure 3. Experimental Results for Learned EM Sensing Applied to an Imaging Task**

For three different poses, we display the images reconstructed with different numbers of coding patterns of the metasurface, M, for the case of only optimizing $\mathbf{\Phi}$ (first row) or jointly optimizing $\mathcal{C}$ and $\mathbf{\Phi}$ (second row). Remarkable improvements of the image quality are achieved by using the joint optimization of $\mathcal{C}$ and $\mathbf{\Phi}$, compared with the case of only optimizing $\mathbf{\Phi}$, when the number of coding patterns of metasurface is less than 9. In this set of experiments, the random initialization is used.

training, except PCA-based $\mathcal{C}$, we have conducted 500 realizations to remove any sensitivity to the choices of random initializations made for m-ANN and r-ANN. Examples of the corresponding optimized coding patterns of the metasurface are displayed in Figure S5.

Figure 3 reports several selected image reconstruction results of the test person with different body gestures using the aforementioned sensing methods. The corresponding coding patterns of the metasurface are displayed in Figure S5. In line with del Hougne et al.,[22] we observe that the sensing quality (here

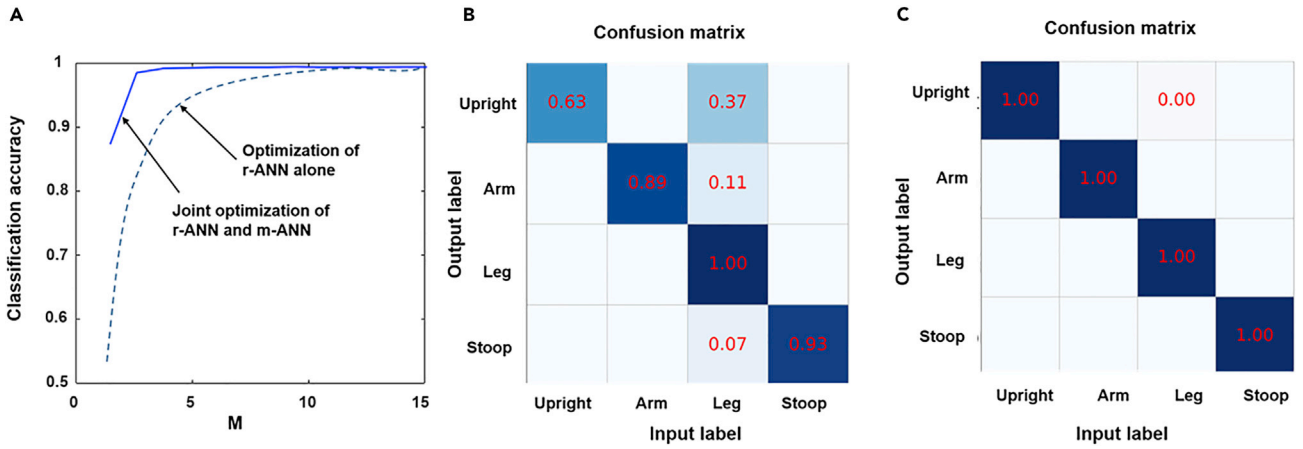



**Figure 4. Experimental Results for Learned EM Sensing Applied to a Gesture-Recognition Task**

Rather than imaging the scene, we now directly process the raw data with the r-ANN in order to classify the gesture displayed in the scene. The metasurface coding patterns $\mathcal{C}$ are initialized randomly.

(A) Dependence of the classification accuracy on the number of coding patterns of the programmable metasurface.

(B) The classification confusion matrix for M = 3 if only $\mathbf{\Phi}$ is optimized.

(C) The classification confusion matrix for M = 3 for jointly optimized $\mathcal{C}$ and $\mathbf{\Phi}$.

image quality) achieved by jointly optimizing physical ($\mathcal{C}$) and digital ($\mathbf{\Phi}$) parameters is significantly better than if only $\mathbf{\Phi}$ is optimized. This may be intuitively expected, since more trainable parameters are available and all a priori knowledge is used in the learned sensing scheme. Unlike Vedula et al.,[28] we do not observe a significant performance dependence on the initialization of the m-ANN (PCA-based or random) in this work.

Our experimental results demonstrate, in line with del Hougne et al.,[22] that simultaneous learning of measurement and reconstruction settings is remarkably superior to the conventional sensing strategies whereby measurement and/or reconstruction are optimized separately (if optimized at all). The benefits of learned sensing are especially strong when the number of measurements is highly limited such that learned sensing enables a remarkable dimensionality reduction. Ultimately, these superior characteristics are enabled by training a unique integrated sensing chain, making use of all available a priori knowledge about the probed scene, task, and constraints on measurement setting and processing pipeline.

### *In Situ* Recognition of Body Gestures

Finally, we consider the task of recognizing body gestures from raw measurement data, i.e., without an intermediate imaging step. Similar to before, m-ANN and r-ANN are merged into a unique sensing chain, which is simultaneously trained using a standard supervised-learning technique, this time to maximize the entire system's gesture classification accuracy. We use the same training and test dataset of four different body gestures as well as the same two-stage training procedure as before. $\mathcal{C}$ is initialized randomly.

Figure 4A reports the average classification accuracy as a function of the number of measurements M. The presented results have been obtained over four different body gestures, and each gesture with 100 test samples. Additionally, the corresponding classification matrices for different sensing methods are reported in Figures 4B and 4C. A classification matrix shows how often a given body gesture is mistakenly classified as one of the other gesture classes. Note that the strong diagonal elements (corresponding to correctly identified body gestures) reflect the achieved recognition accuracy of almost 100% on average by using the learned sensing method. The performance using the learned sensing scheme saturates around M = 5 at 98%, meaning that we extract all task-relevant information with only three coding patterns of the metasurface. For $M \leq 5$, our scheme yields gains in accuracy of the order of 5%–35%, a substantial improvement in the context of classification tasks. The reduction in the number of necessary measurements thanks to the learned sensing scheme will reduce the time needed to acquire and process data, as well as the sensor's energy consumption.

## DISCUSSION

We have reported an experimental implementation of the recently proposed learned EM sensing paradigm applied to two real-life tasks: imaging and gesture recognition. We leveraged a programmable metasurface on the physical layer to shape scene illuminations. Our integrated sensing chain is composed of two dedicated deep ANNs, the m-ANN for smart data acquisition and the r-ANN for instant data processing. Using a VAE formalism, we jointly trained the learnable weights of the measurement process and those of the processing layer to learn an optimal sensing strategy. Thereby, we make use of all available a priori knowledge about the probed scene, the specific sensing task, and the constraints on the measurement setting and processing pipeline. As a result, the learned sensing strategy with simultaneous learning of measurement and reconstruction procedures yields a superior performance compared with conventional sensing strategies that optimize measurement and processing separately (or not at all). The performance improvements are particularly large when the number of measurements is very limited, as we have experimentally demonstrated.

To summarize, we have reported experimental results with immediate practical impact showing how to merge artificial materials and artificial intelligence in the design of a learned EM sensing architecture. Our work paves the path to low-latency microwave sensing for security screening, human-computer interaction, health care, and automotive radar.

## EXPERIMENTAL PROCEDURES

### Design of Programmable Coding Metasurface

The programmable metasurface[29–31] is an ultrathin planar array of electronically reconfigurable meta-atoms. Thanks to its unique capabilities to manipulate EM wave fields in a reprogrammable manner, it has elicited many exciting physical phenomena[39,40] and versatile functional devices,[41] such as computational imagers and sensors,[11–13,20,22,32,42] programmable holography,[43] wireless communication in programmable environments,[44,45] and even analog wave-based computing.[46]

The designed programmable metasurface consists of 32 × 24 meta-atoms operating around 2.4 GHz, as shown in Figure S1, in which the meta-atom with size of 54 × 54 mm is detailed. The designed meta-atom is composed of two substrate layers: the top layer is F4B with a relative permittivity of 2.55 and loss tangent of 0.0019; the bottom layer is FR4 with dimensions of 0.54 × 0.54 $mm^2$. The top square metal patch of size 0.37 × 0.37 $mm^2$ contains a PIN diode (SMP1345-079LF) to control the EM reflection phase of the meta-atom. In addition, a Murata LQW04AN10NH00 inductor with inductance of 33 nH is used to achieve good isolation between the radiofrequency and direct current signals. In our design, the entire programmable metasurface is composed of 3 × 3 metasurface panels, and each panel is composed of 8 × 8 electronically controllable digital meta-atoms. Each metasurface panel is equipped with eight 8-bit shift registers (SN74LV595APW), and eight PIN diodes are sequentially controlled. The adopted clock (CLK) rate is 50 MHz, and the ideal switching time of the PIN diodes is 10 μs. In our specific implementation, the switching time of the coding pattern of the metasurface is set to 20 μs; thus, the time of data acquisition is on the order of M × 20 μs, where M is the total number of the coding patterns.

### Model of the m-ANN

Rather than attempting to write down an analytical forward model of the measurement process as done by del Hougne et al.,[22] we opt for learning a forward model, m-ANN, that automatically accounts for metasurface fabrication inaccuracies and stray reflections. Note that this approach could also readily be applied to alternative hardware capable of generating programmable scene illuminations such as dynamic metasurface antennas[47] or even traditional antenna arrays.[48] Unlike conventional end-to-end deep ANNs that have two ports, our m-ANN has three ports: one inputs the latent variable $\boldsymbol{x}$ of the probed scene (in our work an image or a gesture class), one inputs the metasurface coding pattern $\mathcal{C}$, and the third one outputs the raw measurement $\boldsymbol{y}$ (in our work the raw microwave data), as illustrated in Figure 1. From a deterministic standpoint, we can describe the three-port m-ANN with a pair of coupled nonlinear equations:

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}; \mathcal{W}_{\mathcal{C}}) + \boldsymbol{n} \quad \text{(Equation 2)}$$

$$\text{and } \mathcal{W}_{\mathcal{C}} = \boldsymbol{g}(\mathcal{C}; \boldsymbol{\Theta}) + \boldsymbol{N} \quad \text{(Equation 3)}$$

The variables $\boldsymbol{n}$ and $\boldsymbol{N}$ in Equations 2 and 3 represent the modeling errors, and their entries are considered to be independent identically distributed complex-valued Gaussian random numbers. In Equation 2, $\mathcal{W}_{\mathcal{C}}$ encapsulates all learnable weights of an end-to-end ANN relating the input $\boldsymbol{x}$ to the desired measurement $\boldsymbol{y}$. $\mathcal{W}_{\mathcal{C}}$ imposed with the subscript $\mathcal{C}$ highlights the fundamental fact that these trainable network weights depend on the coding patterns of metasurface $\mathcal{C}$ through Equation 3. In our implementation, these two nonlinear equations, i.e., $\boldsymbol{f}$ and $\boldsymbol{g}$, are modeled with a deep convolutional neural network (CNN) with trainable weights $\mathcal{W}_{\mathcal{C}}$ and $\boldsymbol{\Theta}$, respectively. In this work, we explore the deep residual CNN architectures developed by Li et al.[37] but with eight layers for $\boldsymbol{f}$ and three layers for $\boldsymbol{g}$. Both $\mathcal{W}_{\mathcal{C}}$ and $\boldsymbol{\Theta}$ can be readily learned with a standard supervised-learning procedure from triplet training samples $\{\boldsymbol{x}^{(i)}, \mathcal{C}^{(i)}, \boldsymbol{y}^{(i)}, i = 1, 2, \ldots\}$, where the superscript denotes the index of training samples. More details can be found in Note S4.

### Training of m-ANN and r-ANN

The training of the complex-valued weights of m-ANN and r-ANN is performed using the ADAM optimization method.[49] In addition, we take the deep residual CNN architectures developed by Li et al.[32] for the r-ANN. The complex-valued weights are initialized by random weights with a zero-mean Gaussian distribution of standard deviation $10^{-3}$. The training is performed on a workstation with an Intel Xeon E5-1620v2 central processing unit, NVIDIA GeForce GTX 1080Ti, and 128 GB access memory. The ML platform Tensor Flow[50] is used to design and train the networks in the learned EM sensing system. It will take about 267 h to train the whole learnable sensing pipeline including the m-ANN and r-ANN. Once the r-ANN is well trained, its calculation costs less than 0.6 ms.

### Configuration of Proof-of-Concept Sensing System

The experimental setup, sketched in Figure 1A, consists of a transmitting (TX) horn antenna, a receiving (RX) horn antenna, a large-aperture programmable metasurface, and a vector network analyzer (VNA; Agilent E5071C). The two horn antennas are connected to two ports of the VNA via two 4-m-long 50-Ω coaxial cables, and the VNA is used to acquire the response data by measuring transmission coefficients ($S_{21}$). To suppress the measurement noise level, the average number and filtering bandwidth of the VNA are set to 10 and 10 kHz, respectively. More details can be found in Note S1.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.patter.2020.100006.

## ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China (2017YFA0700201, 2017YFA0700202, and 2017YFA0700203) and in part from the National Natural Science Foundation of China (61471006, 61631007, and 61571117). P.d.H. acknowledges fruitful discussions with the other authors of del Hougne et al.[22]

## AUTHOR CONTRIBUTIONS

L.L. conceived the idea, conducted the theoretical analysis, and wrote the manuscript. L.L. and P.d.H. contributed to conceptualization and write-up of the project. H.-Y.L. conducted experiments and data processing. All authors participated in the experiments and data analysis, and read the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: December 4, 2019
Revised: January 13, 2020
Accepted: February 13, 2020
Published: April 10, 2020

## REFERENCES

1. Semenov, S., Kellam, J., Nair, B., Williams, B., Quinn, M., Sizov, Y., Nazarov, A., and Pavlovsky, A. (2011). Microwave tomography of extremities: 2. Functional fused imaging of flow reduction and simulated compartment syndrome. Phys. Med. Biol. *56*, 2019–2030.
2. Poplack, S.P., Tosteson, T.D., Wells, W.A., Pogue, B.W., and Paulsen, K.D. (2007). Electromagnetic breast imaging: results of a pilot study in women with abnormal mammograms. Radiology *243*, 350–359.
3. Hassan, A.M., and El-Shenawee, M. (2011). Review of electromagnetic techniques for breast cancer detection. IEEE Rev. Biomed. Eng. *4*, 103–118.

4. Mercuri, M., Lorato, I.R., Liu, Y., Wieringa, F., Hoof, C.V., and Torfs, T. (2019). Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor. Nat. Electron. *2*, 252–262.

5. Accardo, J., and Chaudhry, M.A. (2014). Radiation exposure and privacy concerns surrounding full-body scanners in airports. J. Radiat. Res. Appl. Sci. *7*, 198–200.

6. Gonzalez-Valdes, B., Alvarez, Y., Mantzavinos, S., Rappaport, C.M., Las-Heras, F., and Martinez-Lorenzo, J.A. (2016). Improving security screening: a comparison of multistatic radar configurations for human body imaging. IEEE Antennas Propag. Mag. *58*, 35–47.

7. Nan, H., Liu, S., Buckmaster, J.G., and Arbabian, A. (2019). Beamforming microwave-induced thermoacoustic imaging for screening applications. IEEE Trans. Microw. Theory Tech. *67*, 464–474.

8. Li, C., Wang, C., Wei, Y., and Lin, Y. (2019). China's present and future lunar exploration program. Science *365*, 238–239.

9. Orosei, R., Lauro, S.E., Pettinelli, E., Cicchetti, A., Cordini, M., Cosciotti, B., Di Paolo, F., Flamini, E., Mattei, E., Pajola, M., et al. (2018). Radar evidence of subglacial liquid water on Mars. Science *361*, 490–493.

10. Picardi, G. (2005). Radar soundings of the subsurface of mars. Science *310*, 1925–1928.

11. Hunt, J., Driscoll, T., Mrozack, A., Lipworth, G., Reynolds, M., Brady, D., and Smith, D.R. (2013). Metamaterial apertures for computational imaging. Science *339*, 310–313.

12. Fromenteze, T., Yurduseven, O., Imani, M.F., Gollub, J., Decroze, C., Carsenat, D., and Smith, D. (2015). Computational imaging using a mode-mixing cavity at microwave frequencies. Appl. Phys. Lett. *106*, 194104.

13. Sleasman, T., Imani, F.M., Gollub, J.N., and Smith, D.R. (2015). Dynamic metamaterial aperture for microwave imaging. Appl. Phys. Lett. *107*, 204104.

14. Sleasman, T., Imani, M.F., Gollub, J.N., and Smith, D.R. (2016). Microwave imaging using a disordered cavity with a dynamically tunable impedance surface. Phys. Rev. Appl. *6*, 054019.

15. Donoho, D.L. (2006). Compressed sensing. IEEE Trans. Inf. Theory *52*, 1289–1306.

16. Jin, K.H., McCann, M.T., Froustey, E., and Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. *26*, 4509–4522.

17. Li, L., Wang, L.G., Teixeira, F.L., Liu, C., Nehorai, A., and Cui, T.J. (2019). DeepNIS: deep neural network for nonlinear electromagnetic inverse scattering. IEEE Trans. Antennas Propag. *67*, 1819–1825.

18. Li, L., Wang, L.G., and Teixeira, F.L. (2019). Performance analysis and dynamic evolution of deep convolutional neural network for electromagnetic inverse scattering. IEEE Antennas Wirel. Proprag. Lett. *18*, 2259–2263.

19. Liang, M., Li, Y., Meng, H., Neifeld, M.A., and Xin, H. (2015). Reconfigurable array design to realize principal component analysis (PCA)-Based microwave compressive sensing imaging system. IEEE Antennas Wirel. Proprag. Lett. *14*, 1039–1042.

20. Li, L., Ruan, H., Liu, C., Li, Y., Shuang, Y., Alù, A., Qiu, C., and Cui, T.J. (2019). Machine-learning reprogrammable metasurface imager. Nat. Commun. *10*, 1082.

21. Horstmeyer, R., Chen, R.Y., Kappes, B., and Judkewitz, B. (2017). Convolutional neural networks that teach microscopes how to image. https://arxiv.org/abs/1709.07223.

22. del Hougne, P., Imani, M.F., Diebold, A.V., Horstmeyer, R., and Smith, D.R. (2019). Learned integrated sensing pipeline: reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network. Adv. Sci. 1901913, https://doi.org/10.1002/advs.201901913.

23. Chakrabarti, A. (2016). Learning sensor multiplexing design through back-propagation. In In Proceedings of the 30th International Conference on Neural Information Processing Systems, D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, and R. Garnett, eds. (Curran Associates), pp. 3081–3089.

24. Kellman, M.R., Bostan, E., Repina, N.A., and Waller, L. (2019). Physics-based learned design: optimized coded-illumination for quantitative phase imaging. IEEE Trans. Comput. Imaging *5*, 344–353.

25. Vincent, S., Steven, D., Yifan, P., Xiong, D., Stephen, B., Wolfgang, H., Felxi, H., and Goedon, W. (2018). End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. ACM Trans. Graph. *37*, 1–13.

26. Chang, J., Sitzmann, V., Dun, X., Heidrich, W., and Wetzstein, G. (2018). Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. Sci. Rep. *8*, 12324.

27. Muthumbi, A., Chaware, A., Kim, K., Zhou, K.C., Konda, P.C., Chen, R., Judkewitz, B., Erdmann, A., Kappes, B., and Horstmeyer, R. (2019). Learned sensing: jointly optimized microscope hardware for accurate image classification. Biomed. Opt. Express *10*, 6351–6369.

28. Vedula, S., Senouf, O., Zurakhov, G., Bronstein, A., Michailovich, O., and Zibulevsky, M. (2019). Learning beamforming in ultrasound imaging. Proc. Mach. Learn. Res. *102*, 493–511.

29. Sievenpiper, D., Schaffner, J., Loo, R., Tangonan, G., Ontiveros, S., and Harold, R. (2002). A tunable impedance surface performing as a reconfigurable beam steering reflector. IEEE Trans. Antennas Propag. *50*, 384–390.

30. Cui, T.J., Qi, M.Q., Wan, X., Zhao, J., and Cheng, Q. (2014). Coding metamaterials, digital metamaterials and programmable metamaterials. Light Sci. Appl. *3*, e218.

31. Li, L., and Cui, T.J. (2019). Information metamaterials—from effective media to real-time information processing systems. Nanophotonics *8*, 703–724.

32. Li, L., Shuang, Y., Ma, Q., Li, H., Zhao, H., Wei, M., Liu, C., Hao, C., Qiu, C., and Cui, T.J. (2019). Intelligent metasurface imager and recognizer. Light Sci. Appl. *8*, 97.

33. Kingma, D.P., and Welling, M. (2014). Auto-encoding variational Bayes. https://arxiv.org/abs/1312.6114.

34. Doersch, C. (2016). Tutorial on variational autoencoders. https://arxiv.org/abs/1606.05908.

35. Mehta, P., Bukov, M., Wang, C.H., Day, A., Richardson, C., Fisher, C., and Schwab, D.J. (2019). A high-bias, low-variance introduction to machine learning for physicists. Phys. Rep. *810*, 1–124.

36. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. Proc. IEEE *86*, 2278–2324.

37. Li, L., Jafarpour, B., and Mohammad-Khaninezhad, M.R. (2013). A simultaneous perturbation stochastic approximation algorithm for coupled well placement and control optimization under geologic uncertainty. Comput. Geosci. *17*, 167–188.

38. del Hougne, P., Imani, M.F., Fink, M., Smith, D.R., and Lerosey, G. (2018). Precise localization of multiple noncooperative objects in a disordered cavity by wave front shaping. Phys. Rev. Lett. *121*, 063901.

39. Zhang, L., Chen, X.Q., Liu, S., Zhang, Q., Zhao, J., Dai, J.Y., Bai, G.D., Wan, X., Cheng, Q., Castaldi, G., et al. (2018). Space-time-coding digital metasurfaces. Nat. Commun. *9*, 4334.

40. Zhang, L., Chen, X.Q., Shao, R.W., Dai, J.Y., and Cui, T.J. (2019). Breaking reciprocity with space-time-coding digital metasurfaces. Adv. Mater. *31*, 1904069.

41. Yang, H., Cao, X., Yang, F., Gao, J., Xu, S., Li, M., Chen, X., Zhao, Y., Zheng, Y., and Li, S. (2016). A programmable metasurface with dynamic polarization, scattering and focusing control. Sci. Rep. *6*, 35692.

42. Gollub, J.N., Yurduseven, O., Trofatter, K.P., Arnitz, D., Imani, M.F., Sleasman, T., Boyarsky, M., Rose, A., Pedross-Engel, A., et al. (2017). Large metasurface aperture for millimeter wave computational imaging at the human-scale. Sci. Rep. *7*, 42650.

43. Li, L., Jun Cui, T., Ji, W., Liu, S., Ding, J., Wan, X., Bo Li, Y., Jiang, M., Qiu, C., and Zhang, S. (2017). Electromagnetic reprogrammable coding-metasurface holograms. Nat. Commun. *8*, 197.

44. del Hougne, P., Fink, M., and Lerosey, G. (2019). Optimally diverse communication channels in disordered environments with tuned randomness. Nat. Electron. *2*, 36–41.

45. Zhao, J., Yang, X., Dai, J.Y., Cheng, Q., Li, X., Qi, N.H., Ke, J.C., Bai, G.D., Liu, S., Jin, S., et al. (2019). Programmable time-domain digital-coding metasurface for non-linear harmonic manipulation and new wireless communication systems. Natl. Sci. Rev. *6*, 231–238.

46. del Hougne, P., and Lerosey, G. (2018). Leveraging chaos for wave-based analog computation: demonstration with indoor wireless communication signals. Phys. Rev. X *8*, 041037.

47. Sleasman, T., Imani, M.F., Diebold, A.V., Boyarsky, M., Trofatter, K.P., and Smith, D.R. (2019). Implementation and characterization of a two-dimensional printed circuit dynamic metasurface aperture for computational microwave imaging. https://arxiv.org/abs/1911.08952.

48. Fenn, A.J., Temme, D.H., Delaney, W.P., and Courtney, W.E. (2000). The development of phased-array radar technology. Lincoln Lab. J. *12*, 20.

49. Kingma, D.P., and Ba, J.L. (2014). Adam: a method for stochastic optimization. https://arxiv.org/abs/1412.6980.

50. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: a system for large-scale machine learning. 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.

**PATTER, Volume *1***

## Supplemental Information

# Intelligent Electromagnetic Sensing with Learnable Data Acquisition and Processing

**Hao-Yang Li, Han-Ting Zhao, Meng-Lin Wei, Heng-Xin Ruan, Ya Shuang, Tie Jun Cui, Philipp del Hougne, and Lianlin Li**

## Supplementary Note 1. Sensing system based on the programmable metasurface

The configuration of proposed intelligent sensing system, with reference to **Figure 1a** in main text, consists of a pair of horn antennas, a vector network analyzer (VNA, Aligent E5071C), and a large-aperture programmable metasurface. The operational principle of the presented sensing system is described as following. Antenna 1, connected to port-1 of VNA, is used to emit periodically microwave illumination signals, which are shaped by the m-ANN-driven programmable metasurface. After being scattered by the subject of interest, the wavefield shaped by the metasurface are received by Antenna-2 connected to port-2 of VNA. Finally, the received microwave raw data are instantly processed by the r-ANN, producing the desired imaging or recognition results.

**Figures S1a** show the photos of the front and back views of the designed programmable metasurface. For the sake of fabrication limitation, the whole metasurface is designed to be composed of $3\times 4$ identical metasurface panels, and each panel has $8 \times 8$ meta-atoms. Each meta-atom has a size of $54\times54\text{mm}^2$, thus the whole metasurface has a size of $1.7\times1.3\text{m}^2$ in total. The metasurface is electronically controlled with a FPGA-based Micro-Control-Unit (MCU), as shown in the insert of **Figure S1a**. In addition, the geometrical parameters of the electronically-controllable digital meta-atom are detailed in **Figure S1b.** From **Figure S1c,** each metasurface panel is equipped with eight 8-bit shift registers (SN74LV595APW), and every 8 PIN diodes share a shift register. With the use of shift registers, 8 PIN diodes are sequentially controlled. The MCU sends the commands over 24 independent branch channels, leading to real-time manipulations of all PIN diodes. The MCU works with one common clock (CLK) signal. In our work, the adopted CLK is 50MHz, and the switching time of PIN diode is about 10us each cycle. We remark that the control strategy can be extended for more PIN diodes by concatenating more metasurface

panels in a straightforward manner, allowing adjustable rearrangement of metasurface panels for various application needs.

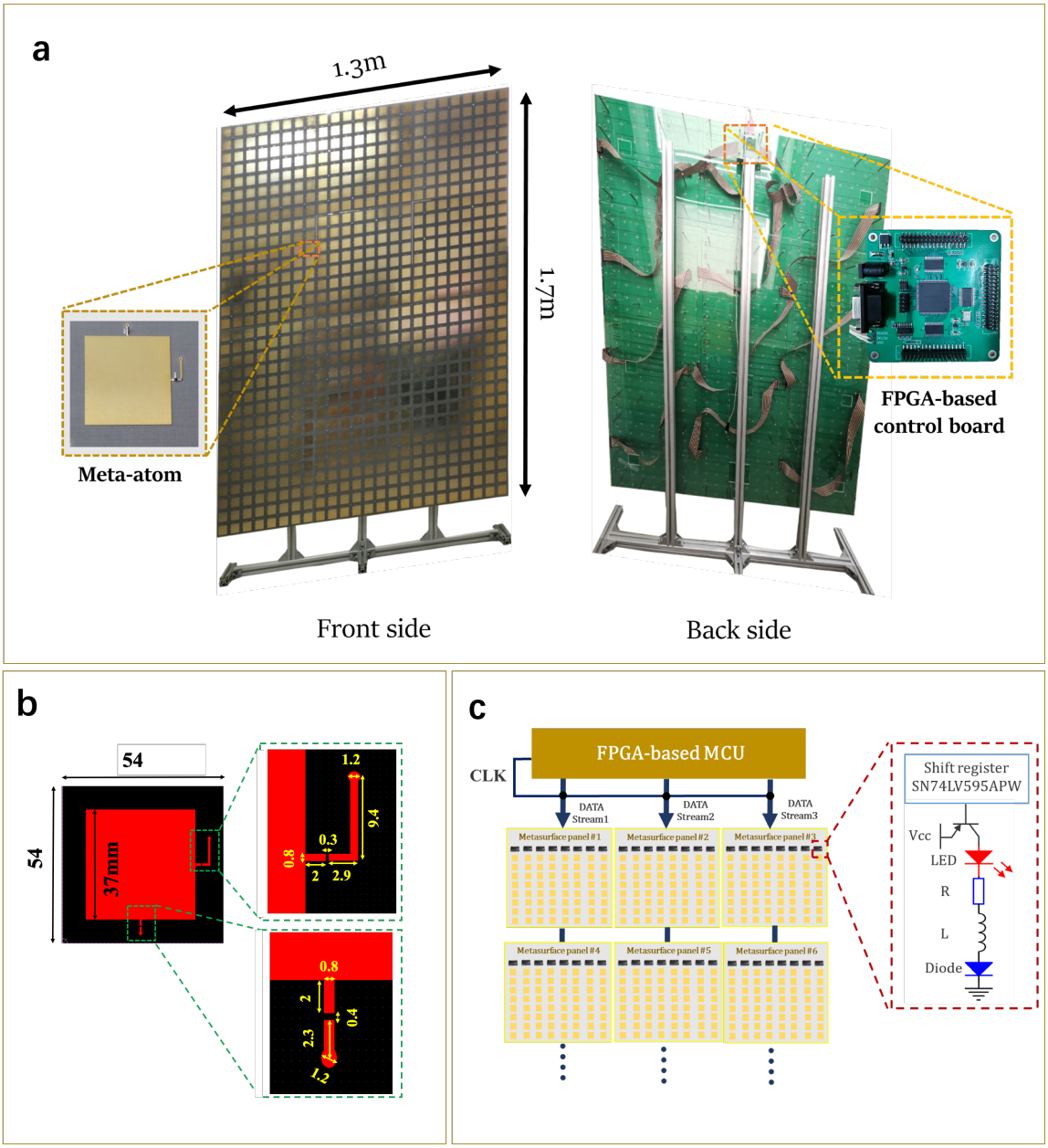



**Figure S1**. **(a),** the photos of front-side and back-side of the designed programmable metasurface. Here, the FPGA-based micro control unit (MCU) has been inserted as well in the back-side view picture. **(b)** the geometrical parameters of the designed electronically-controllable meta-atom of the metasurface. **(c)** the illustrative control strategy of the programmable metasureface.

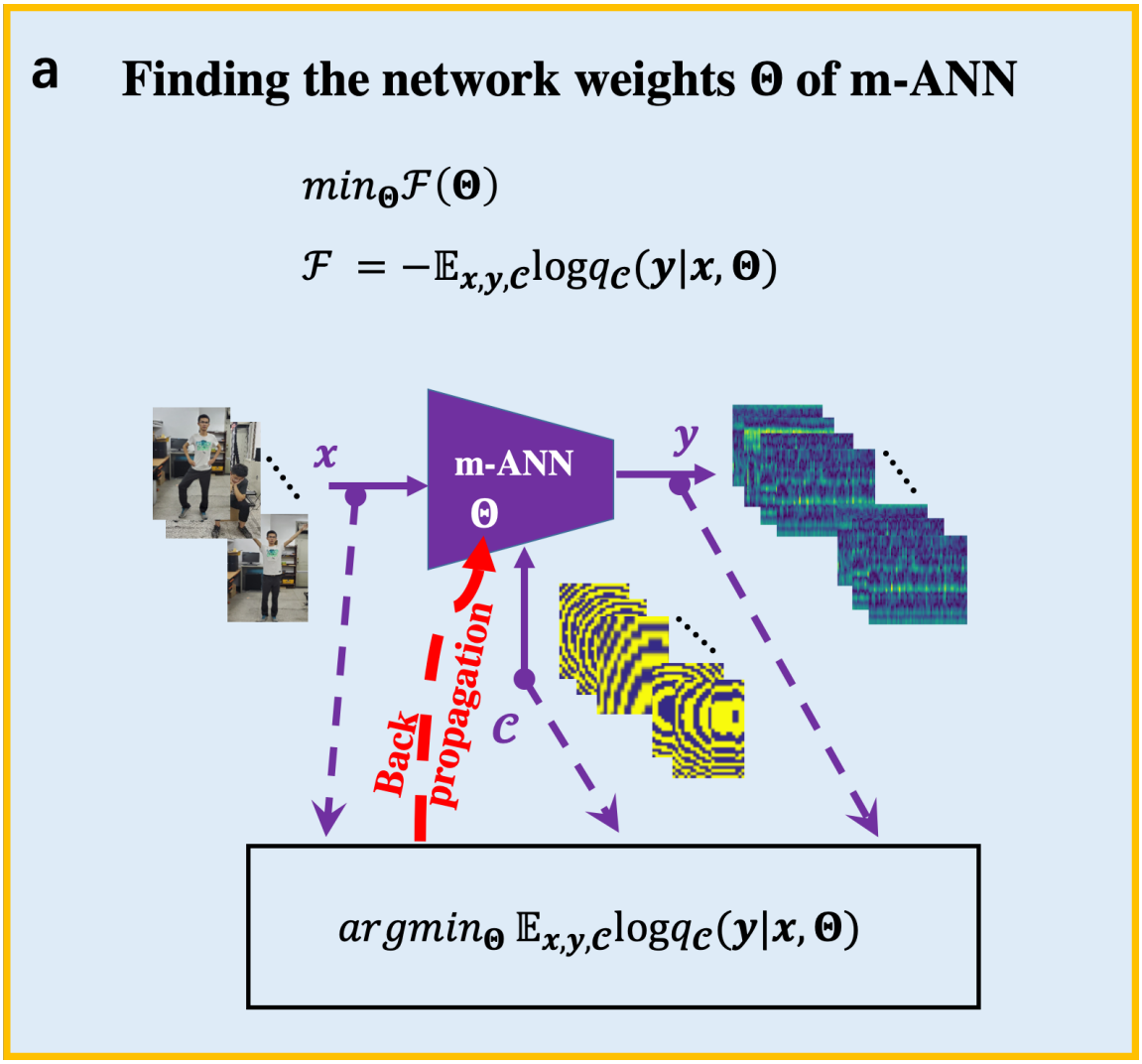
## Supplementary Note 2. Discussions about the VAE objective function

In this note, we would like to discuss briefly Eq. (1) outlined in main text, and elaborate on associated optimization algorithm. Generally speaking, the measurement procedure of the proposed intelligent sensing strategy can be viewed as an end-to-end process that given a scene $\boldsymbol{x}$ (image or class label of probed subject) generates a set of measurements $\boldsymbol{y}$ by sampling from a $\boldsymbol{\mathcal{C}}$-controllable conditional distribution $\boldsymbol{y}\sim q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})$, where $\boldsymbol{\mathcal{C}}$ encapsulates all trainable parameters of the hardware setting, i.e., the user-controlled coding pattern of metasurface in our work. This conditional distribution is known as the likelihood in the framework of Bayesian analysis, and can be understood as a stochastic measurement model. Basically, the goal of data processing pipeline is to produce an estimate $\hat{\boldsymbol{x}}$ of the scene $\boldsymbol{x}$ given the measurements $\boldsymbol{y}$. Basically, the estimator $\hat{\boldsymbol{x}}$ serves as an inverse action of the measurement process, and can be realized with a deep ANN with network weights $\boldsymbol{\Phi}$. Similar to the measurement process, we denote the estimator with a parametric conditional distribution $\hat{\boldsymbol{x}}\sim p(\hat{\boldsymbol{x}}|\boldsymbol{y},\boldsymbol{\Phi})$. We propose to simultaneously learn the learnable parameters, i.e., $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\Phi}$, of both the measurement process and the reconstruction operator in context of VAE, such as to optimize the whole sensing performance in a specific task. In light of VAE, the optimal choices of $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\Phi}$ can be achieved by minimizing the following objective function, i.e.,

$$\mathcal{L}(\boldsymbol{\mathcal{C}},\boldsymbol{\Phi}) = -\mathbb{E}_{q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})}[\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi})] + \mathrm{KL}(q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})||p(\boldsymbol{y})) \tag{S1}$$

Note that the mathematic expectation in the first term of the right hand of Eq. (S1) is taken over the distribution $q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})$, which embodies the $\boldsymbol{\mathcal{C}}$-controllable measurements and the reconstruction network as a whole. In a nutshell, this term serves as a likelihood, which is used to measure the reconstruction error over $q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})$. In contrast, the second term is characterized by KL-divergence, which, as a regu larizer, encourages the measurement distribution $q_{\boldsymbol{\mathcal{C}}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})$ to be close to a chosen prior $p(\boldsymbol{y})$. Here, $p(\boldsymbol{y})$ is chosen to be

zero-mean Gaussian distribution with maximum Shannon information entropy. As such, each measurement is optimized to capture as much information of the probed scene as possible.



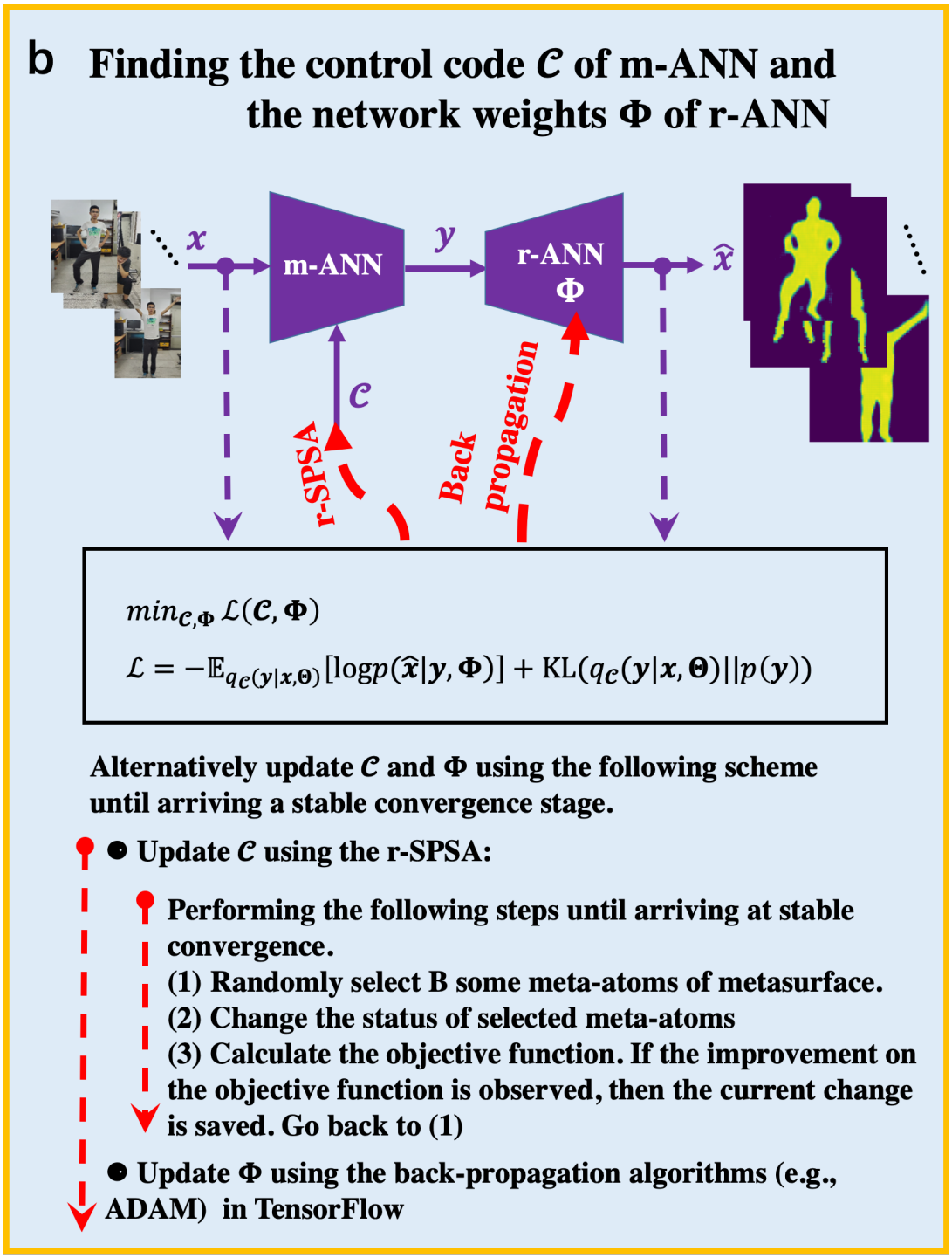



**Figure S2**. (a) the flow chart of learning the network weights of the m-ANN. (b) the flow chart of determining the optimal settings of the coding pattern of the metasurface and the network weights of the r-ANN.

In numerical implementation, the expectation term $-\mathbb{E}_{q_{\mathcal{C}}(y|x,\Theta)}[\log p(x|y,\Phi)]$ is

replaced by finite-sample statistical mean approximation over the training dataset. As for $\mathrm{KL}(q_{\boldsymbol{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})||p(\boldsymbol{y}))$, it can be analytically treated under the Gaussian assumption, as detailed in ref. 1 in main text. In addition, in our implementation, Eq. (S1) is slightly modified as following, i.e.,

$$\mathcal{L}(\boldsymbol{C},\boldsymbol{\Phi}) = -\mathbb{E}_{q_{\boldsymbol{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})}[\log p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi})] + \gamma \mathrm{KL}(q_{\boldsymbol{C}}(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\Theta})||p(\boldsymbol{y})) \quad \text{(S2)}$$

Here, $\gamma$ is introduced to tradeoff the contributions from the data misfit and the prior-based regularization.

It is noted that in minimizing Eq. (S1), it involves two sets of different optimization variables, i.e., the continuously adjustable weights $\boldsymbol{\Phi}$ in the r-ANN and $\boldsymbol{\Theta}$ in the m-ANN, and the binary controllable variables $\boldsymbol{\Theta}$ in the m-ANN. Apparently, the optimization with respect to $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ can be efficiently realized with the well-known back-propagation (BP) algorithm, which can be accomplished with well-developed optimizers in TensorFlow. However, it is really challenging to minimize Eq.(S1) with respect to the binary control coding sequences $\boldsymbol{C}$ since it involves a NP-hard combinatorial optimization problem. To surpass this difficulty, the randomized simultaneous perturbation stochastic approximation (r-SPSA), originally developed for the problem of optimal well place and control in area of petroleum engineering, is slightly modified for our problem. This heuristic optimization approach relies on two randomized descent strategies. First, as done by stochastic gradient descent approach, at each iteration, a fraction of training samples is randomly selected to determine a descent direction. Consequently, the concept of batch size is also applicable. Second, at each iteration, as done by so-called randomized coordinate descent method, only a fraction of optimization components chosen to be updated. Here, partial coding meta-atoms of metasurface are randomly selected, and their binary status are changed to their opposites correspondingly. If the change leads to the improvement on the objective function defined over randomly selected training samples, we save such change and go into next iteration. Otherwise, we need to randomly re-select some of coding meta-atoms of metasurface and perform above operations. Repeat such procedure until some stop criterion

is arrived. More details about the implementation of the proposed r-SPSA optimization algorithm can be seen in **Figure S2**.

## Supplementary Note 3. Experimental settings

In order to evaluate the contribution of the joint training of the measurement setup and data processing pipeline, we have designed a two-stage experiment. First, we train the reconstruction network alone while fix the measurement setting (i.e., the programmable metasurface with pre-specified control coding patterns). Second, we use a pre-convergence checkpoint of the reconstruction network as a starting point for the joint training. At this stage, both the reconstruction network and the measurement setting are jointly trained. In order to factor out the undesired influence of the optimization algorithm on the sensing results, we train the reconstruction network in both stages with the same optimizer (ADAM, batch-size=10, initial learning rate=0.005), and train the measurement setting (i.e., the control coding pattern of metasurface) using the r-SPSA algorithm (batch-size =200, and the number of elements each iteration=20).

For all the experiments throughout the paper, the *learned measurement* refers to the setting where the control coding pattern of programmable metasurface is trained alone and the reconstruction network is fixed; the *learned reconstruction* refers to the setting where the reconstruction network is trained alone, and the control coding pattern of programmable metasurface is fixed; the *learned measurement-reconstruction* refers to the setting where the control coding pattern of metasurface is jointly trained with the reconstruction network.

## Supplementary Note 4. Discussions about the m-ANN

Here, we would like to provide physical insights into the m-ANN by investigating its connection with the classical EM scattering mechanism. To that end, the investigation domain including the subject is uniformly divided into M subgrids, and each subgrid is

occupied by an ideal dipole. In light of coupled dipole method (CDM), when the probed subject is illuminated by an EM wavefield, the resultant scattering electrical wavefield at $\boldsymbol{r}$ outside the investigation domain is governed by the following equations:

$$E^{(sca)}(\boldsymbol{r}) = \sum_{m=1}^{M} \mathrm{G}(\boldsymbol{r}, \boldsymbol{r}_m)\alpha(\boldsymbol{r}_m)E(\boldsymbol{r}_m) \tag{S4}$$

and

$$E(\boldsymbol{r}_n) = E^{(inc)}(\boldsymbol{r}_n) + \sum_{m=1, m\neq n}^{M} \mathrm{G}(\boldsymbol{r}_n, \boldsymbol{r}_m)\alpha(\boldsymbol{r}_m)E(\boldsymbol{r}_m) \tag{S5}$$

$$n = 1,2,\dots,M$$

Herein, $\alpha(\boldsymbol{r}_m)$ represents the polarizability of the equivalent dipole at the location of $\boldsymbol{r}_m$, $E(\boldsymbol{r}_m)$ means the internal electrical field induced at $\boldsymbol{r}_m$, and $\mathrm{G}$ is the Green's function for surrounding background environment. $E^{(inc)}(\boldsymbol{r}_n)$ denotes the illumination radiated from the metasurface controlled with a control coding pattern. Note that we have explicitly made the scalar simplification in Eqs. (S4) and (S5) for the sake of simplicity, however, the methodology presented here can be extended to full-vectorial cases in a straightforward manner.

Now, our primary goal is to calculate $E^{(sca)}(\boldsymbol{r})$ given $\{E^{(inc)}(\boldsymbol{r}_n)\}$ and $\{\alpha(\boldsymbol{r}_m)\}$, where a critical issue is to estimate the intermediate variables $\{E(\boldsymbol{r}_n)\}$ by solving Eq.(S5). Once $\{E(\boldsymbol{r}_n)\}$ are known, the scattering wavefield $E^{(sca)}(\boldsymbol{r})$ can be obtained by using Eq. (S4). However, it is an open challenging issue to efficiently solve Eq. (S5) when the Green's function $\mathrm{G}$ cannot be treated in a tractable way. As a matter of fact, even if the $\mathrm{G}$ can be efficiently treated, it remains not an easy task to solve the large-scale problem of Eq. (S5) from the computational viewpoint, for instance, this inverse problem usually suffers from the notorious ill-posedness in the large-scale case. In the era of AI (Artificial Intelligence), ML techniques, especially deep learning, can be explored to address above difficulties. Particularly, with a standard supervised training procedure, a well-defined end-to-end mapping from $\{E^{(inc)}(\boldsymbol{r}_n)\}$ or $\{\alpha(\boldsymbol{r}_m)\}$ to $E^{(sca)}(\boldsymbol{r})$ can be learned from a number of labeled training data. In this section, we would like to investigate a connection between the solution to Eqs. (S4)-(S5) and the deep ANN, as outlined in **Figure S3**.

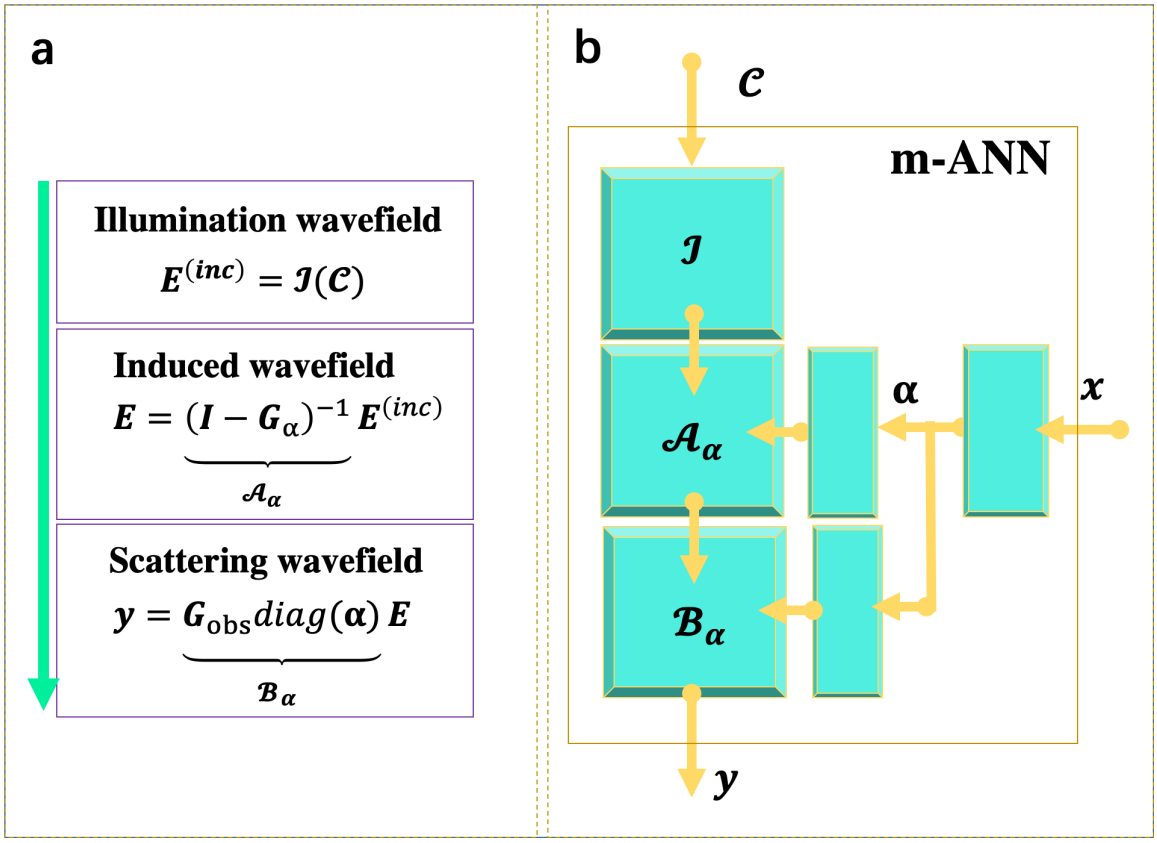



**Figure S3**. The connection between the proposed measurement procedure and the three-port m-ANN. (**a**), the acquired signal can be modelled with three cascaded networks. The first one is characterized by $\boldsymbol{E}^{(inc)} = \boldsymbol{\mathcal{I}}(\boldsymbol{\mathcal{C}})$, which relates nonlinearly the illumination wavefield $\boldsymbol{E}^{(inc)}$ to the coding pattern $\boldsymbol{\mathcal{C}}$ of the metasurface. Other two are α-dependent fully-connected networks which are mathematically characterized by $\boldsymbol{E} = (\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}\boldsymbol{E}^{(inc)} = \mathcal{A}_\alpha \boldsymbol{E}^{(inc)}$ and $\boldsymbol{y} = \boldsymbol{G}_{\text{obs}} diag(\alpha)\boldsymbol{E} = \mathcal{B}_\alpha \boldsymbol{E}$, respectively. Here, $\boldsymbol{E} = \boldsymbol{G}_{\text{obs}} diag(\boldsymbol{\alpha})\boldsymbol{E}$ is a compact representation of **Eq.(S4)**. The dependence on α from these two fully-connected networks are further established with deep ANNs, as demonstrated in **Figure S3b**. (**b**) the deep ANN representation of (a). The polarizability distribution $\boldsymbol{\alpha}$ could be a linear or nonlinear function of the input scene $\boldsymbol{x}$. For instance, when $\boldsymbol{x}$ indicates the class label of the scene, $\boldsymbol{\alpha}$ will be nonlinearly related to $\boldsymbol{x}$ via a deep generative network. Such nonlinear mappings are modelled with deep CNNs in our work. Additionally, the nonlinear operator $\boldsymbol{\mathcal{I}}$ is also approximated by a deep CNN in our implementation.

First, we consider to solve Eq. (S5) by using the direct matrix inversion technique. To this end, we rewrite Eq.(S5) in a compact form as following, i.e.,

$$\boldsymbol{E} = \boldsymbol{E}^{(inc)} + \boldsymbol{G}_\alpha \boldsymbol{E} \tag{S6}$$

Herein, $\boldsymbol{E}$ and $\boldsymbol{E}^{(inc)}$ are M-length column vectors, in which the $n$th elements are $E(\boldsymbol{r}_n)$ and $E^{(inc)}(\boldsymbol{r}_n)$, respectively. $\boldsymbol{G}_\alpha$ denotes a α-dependent matrix with a size of $M \times M$,

whose diagonal entries are zero, and the (*n, m*)-entry is $\mathrm{G}(\boldsymbol{r}_n, \boldsymbol{r}_m)\alpha(\boldsymbol{r}_m)$. Apparently, Eq. (S6) can be solved with a matrix inversion, i.e.,

$$\boldsymbol{E} = (\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}\boldsymbol{E}^{(inc)} \tag{S7}$$

Note that the wavefield $\boldsymbol{E}$ is linearly related to illumination $\boldsymbol{E}^{(inc)}$ through $(\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}$. As argued above, it is more feasible to learn an agent for $(\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}$ with ML techniques than calculate it directly. Furthermore, the matrix $(\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}$ varies with $\alpha$ in a nonlinear way. Here, we would like to emphasize that the illumination wavefield $\boldsymbol{E}^{(inc)}$ depends on the control coding pattern of metasurface $\boldsymbol{C}$, and that the polarizability $\alpha$ characterizes the subject. In order to establish the nonlinear relation mapping from $\alpha$ to $(\boldsymbol{I} - \boldsymbol{G}_\alpha)^{-1}$, a standard ANN can be used as well. In addition, the illumination wavefield $\boldsymbol{E}^{(inc)}$depends nonlinearly on the control coding pattern of metasurface $\boldsymbol{C}$, and such nonlinear operator can be modeled with a standard end-to-end ANN as well. Inspired by these observations, we introduce a three-port ANN for charactering the whole measurement procedure of the proposed sensing system, as outlined in **Figure S3**. The resultant ANN has three ports: one is used to receive the scene properties, i.e., $\alpha(\boldsymbol{r}_m)\}$, one is for the control coding pattern of metasurface $\boldsymbol{C}$, and the other is for outputting the microwave raw data $\boldsymbol{y}$. Such three-port measurement ANN can be learned with a standard supervised training procedure.

## Supplementary Note 5. Comparing with PCA-based measurements

The proposed full-ANN-driven intelligent sensing strategy works in a nonlinear ML way, as opposed to PCA (linear) technique. Here, we would like to highlight two major benefits of our sensing technique over PCA method. First, since the PCA approach requires an analytical measurement model, one may obtain the superior performance of our sensing strategy in any scenario where the PCA approach could be applicable, from the viewpoint of computational efforts. Second, the ability to synthesize the PCA measurement modes is highly dependent on the number of metasurface meta-atoms, implying that the PCA approach is suitable only for scenarios where the number of metasurface meta-atoms is

really large.

In our experiments, the coding patterns of the metasurface are initialized using the PCA method, and then jointly optimized along with the r-ANN. The significant improvement on the objective function can be clearly observed, especially when the number of control coding patterns M is very limited. Correspondingly, the imaging results can be clearly improved with our method, highlighting our method's unique ability to perform sensing tasks with respect to the PCA method. This does make sense since the presented intelligent sensing approach has more controllable degrees of freedom, and accounts for more prior knowledge about the scene, the data measurement and processing into the entire sensing chain, as opposed to the PCA approach which only considers the prior information on the scene.

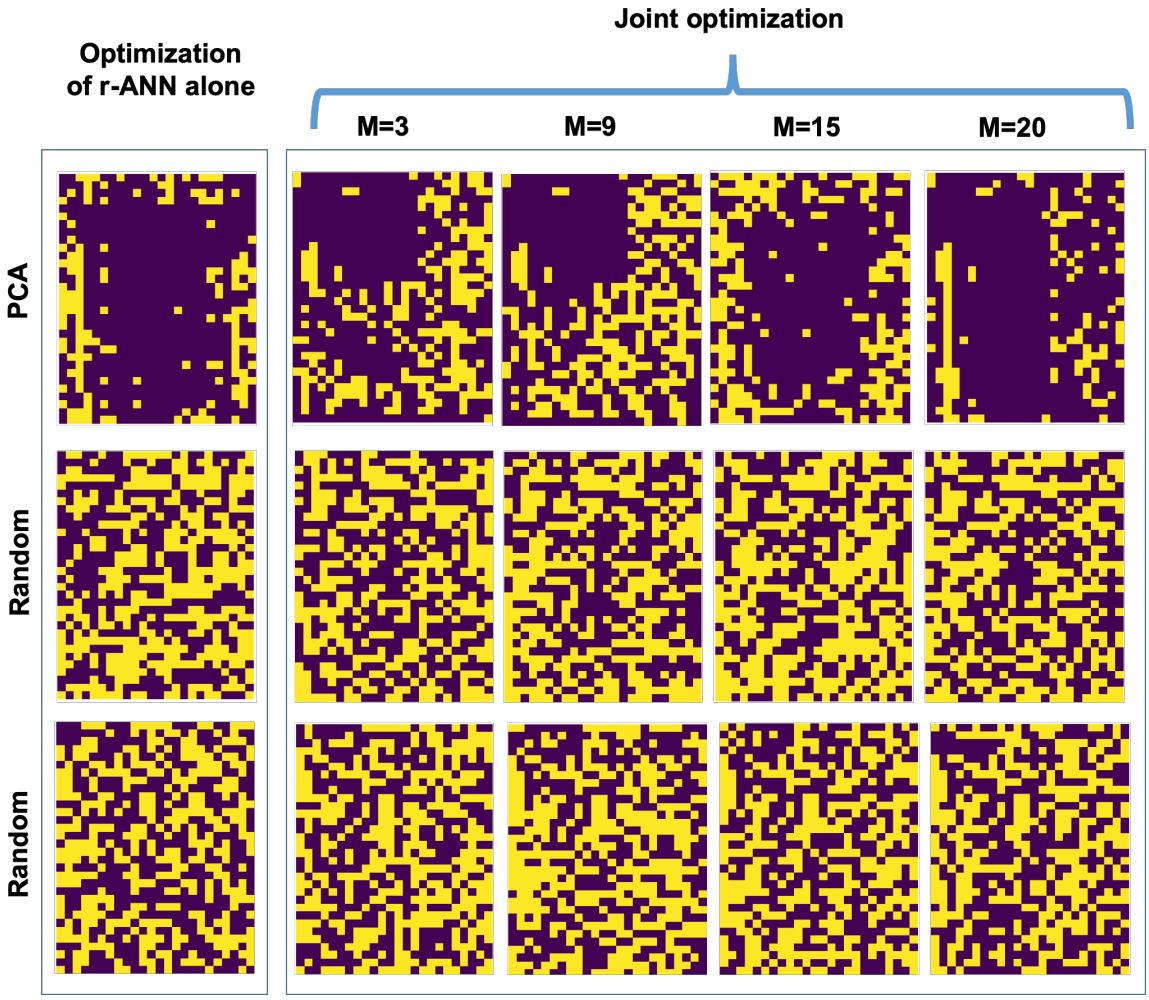



**Figure S4**. Selected optimized coding patterns of the metasurface, corresponding to that involved in Figure 2 in main text. From these figures, it can be observed that the coding patterns of the metasurface are remarkably changed when the measurements are highly limited.

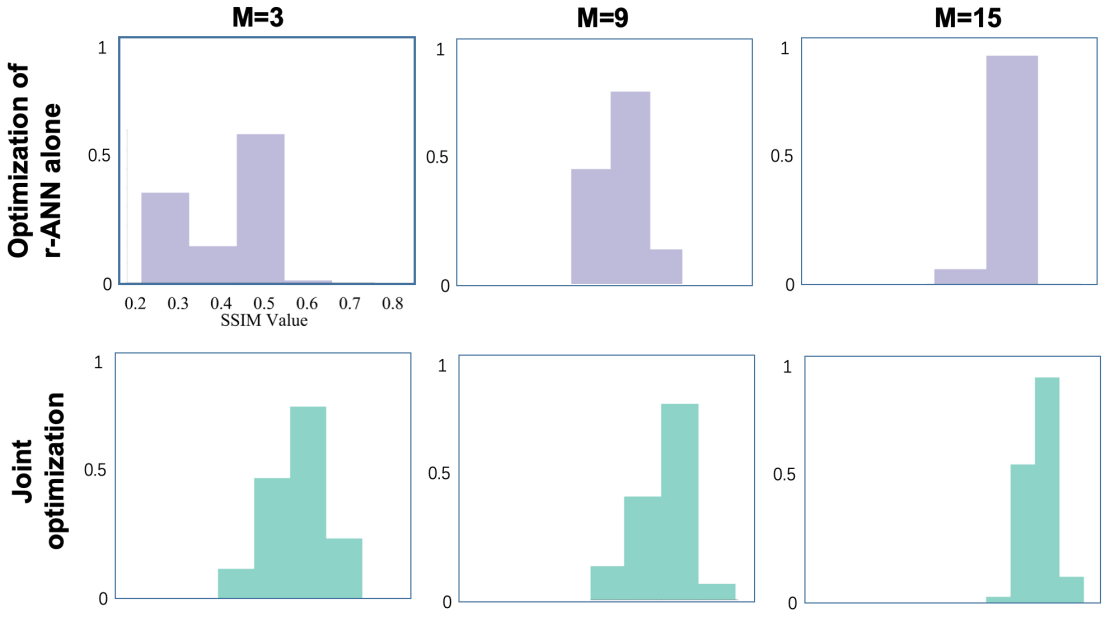



**Figure S5**. The dependences of the SSIMs of the images obtained by different sensing methods on the number of coding patterns M of the programmable metasurface. These results correspond to Figures 2 and 3 in main text.