Supplementary information for

Autonomous aeroamphibious invisibility cloak with stochastic-evolution learning

Chao Qian[1,2,3,†,*], Yuetian Jia[1,2,3,†], Zhedong Wang[1,2,3,†], Jieting Chen[1,2,3], Pujing Lin[1,2,3], Xiaoyue Zhu[1,2,3], Erping Li[1,2,3], and Hongsheng Chen[1,2,3,*]

[1]*ZJU-UIUC Institute, Interdisciplinary Center for Quantum Information, State Key Laboratory of Extreme Photonics and Instrumentation, Zhejiang University, Hangzhou 310027, China.*

[2]*ZJU-Hangzhou Global Science and Technology Innovation Center, Key Lab. of Advanced Micro/Nano Electronic Devices & Smart Systems of Zhejiang, Zhejiang University, Hangzhou 310027, China.*

[3]*Jinhua Institute of Zhejiang University, Zhejiang University, Jinhua 321099, China.*

[†]*These authors contributed equally to this work.*

[*]*Corresponding authors: chaoq@intl.zju.edu.cn (C. Qian); hansomchen@zju.edu.cn (H. Chen)*

Supplementary information guide:

- Supplementary Note 1: Design of the meta-atom
- Supplementary Note 2: Physical principle of the spatiotemporal metasurfaces
- Supplementary Note 3: Comparison between spatial-only and spatiotemporal metasurfaces
- Supplementary Note 4: Broadband generalization of intelligent aeroamphibious cloak
- Supplementary Note 5: Generation-elimination network
- Supplementary Note 6: Supervised learning loss and the accuracies
- Supplementary Note 7: Gyroscope detector
- Supplementary Note 8: Camera and environment discrimination network (EDN)
- Supplementary Note 9: Realization of an intelligent electromagnetic detector
- Supplementary Note 10: Automatic control system of the spatiotemporal metasurfaces
- Supplementary Note 11: Working flowchart of intelligent invisible drone
- Supplementary Note 12: Experiment setup of invisible drone against amphibious background

Other supplementary information includes:

- Supplementary Videos 1 & 2 (.mp4 format). This video shows the experimental setup and results at different outdoor test sites. When the intelligent invisible drone freely flies in the sky and passes through a conical detection region, three receivers record the on-site scattering wave in real time. At the same time, we compare the performance of cloaked drone with bare drone.

Supplementary Note 1: Design of the meta-atom

To realize the spatiotemporal metasurface and obtain the phase state with uniform coverage of 2π, the first step is to design a high-performance programmable metasurface [S1-S3]. The geometrical details of the spatiotemporal meta-atom are illustrated in Fig. S1 and comprise an irregular octagon metal patch and two metal strips patched on the dielectric substrate. The dielectric substrate used in the design is F4B with ε_r = 2.65 and a thickness of 3 mm. The period p of the unit cell is 40 mm. The bottom layer is also a copper patch, serving as the ground to realize the reflective metasurfaces. Two PIN diodes (SMP1320-079L from Skyworks) are placed between the octagon patch and two strips, which act as biasing lines for each diode. A metal via hole with a radius of 0.5 mm is further employed on the octagon patch to connect the top layer with the ground. The commercial software CST Microwave Studio is applied to carry out full-wave simulations and investigate the electromagnetic response of the unit cell. In the simulations of the unit cell, periodic boundary conditions are applied along both *x* and *y* directions, and Floquet ports are used along the -*z* direction. A normally incident plane wave (with *x*-polarized electric field) is assumed to calculate the reflection coefficient of the unit cell under different PIN states. In the optimization stage, we investigate the biasing voltage of two PIN diodes with four states (on-on, on-off, off-on, and off-off) through adjusting l_1, l_2 and l_3 with fixed width $w_1 \sim w_5$. We observe that four distinct reflection responses are achieved with about 90° phase difference at around 3.1 GHz, while the reflection amplitude remains almost unity when l_1 = 21 mm, l_2 = 35 mm and l_3 = 17 mm.

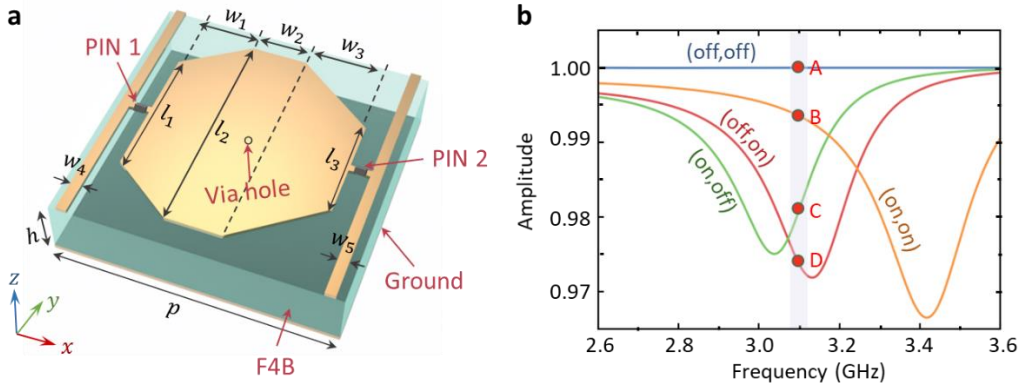

Figure S1 | Three-dimensional (3D) illustration of the tunable meta-atom. a, The parameters are $p = 40,\ h = 3,\ w_1 = 10,\ w_2 = 8,\ w_3 = w_1 + \frac{w_2}{3},\ w_4 = w_5 = 1,\ l_1 = 21,\ l_2 = 35,\ l_3 = 17$ (unit: mm) and the via hole is located at the center of F4B substrate. **b**, Reflected amplitude of the meta-atom by applying different bias voltages across the loaded diodes.

Supplementary Note 2: Physical principle of the spatiotemporal metasurfaces

We consider microwave reconfigurable metasurfaces that incorporate two electronic PIN diodes. By applying different dc voltage, the reflection response of metasurfaces can be switched among four discrete states, (on, on), (on, off), (off, on), and (off, off). We optimize the geometries of metasurfaces to attain an interval of π/2 among four states while keeping the amplitude as high as possible. On this basis, we introduce time-varying modulation into metasurfaces to revamp reconfigurable metasurfaces into spatiotemporal metasurfaces [S4-S7]. Here, we consider periodic time-varying series that consists of *L* segments, and the value of each segment is one of the four discrete reflection states. The reflection state keeps constant in each segment. Mathematically, such time-varying series can be written as,

$$\Gamma(t) = \sum_{l=1}^{L} \Gamma_l G_l(t) \tag{S1}$$

where L is the number of time-varying series, and Γ_l is the reflecting coefficient at lth segment. $G_l(t)$ is the gate function, expressed as,

$$G_l(t) = \begin{cases} 1, & (l-1)T/L \le t < lT/L \\ 0, & else \end{cases} \tag{S2}$$

where T is the period of time-varying series, l is an integer, ranging from 1 to L. Evidently, $G_l(t)$ is a periodic square-wave signal, which is nonzero only in the l^{th} segment. According to Fourier theorem, $\Gamma(t)$ can be decomposed into a sum of a series of orthogonal complex exponential functions with different angular frequencies:

$$\Gamma(t) = \sum_{k=-\infty}^{+\infty} \epsilon_k \, exp(-2\pi i k \Delta f t) \tag{S3}$$

where *k* is the order of the complex exponential term, $\Delta f = 1/T$ is the frequency of the first order complex exponential term, ϵ_k is the Fourier coefficient of the k^{th} complex exponential term. In frequency domain, each complex exponential function is associated with a single frequency harmonic. The frequency of the k^{th} harmonic is $k\Delta f$. And the amplitudes and initial phases of the harmonic is expressed by the Fourier coefficient ϵ_k. The Fourier coefficient ϵ_k can be calculated as,

$$\epsilon_k = \frac{1}{T} \int_0^T \Gamma(t) \, exp(2\pi i k \Delta f t) \, dt = \frac{1}{T} \int_0^T \sum_{l=1}^{L} \Gamma_l G_l(t) \, exp(2\pi i k \Delta f t) \, dt \tag{S4}$$

From the above equation, ϵ_k is determined by the time-varying series. In turn, by suitably designing the time-varying series, we can actively alter the magnitude and phase of a given harmonic wave. By changing the period of time-varying series, the frequency of each harmonic can be varied. We simplify Eq. (S4) as,

$$\epsilon_k = \frac{1}{T}\sum_{l=1}^{L}\Gamma_l \int_0^T G_l(t)\, exp(2\pi i k \Delta f t)\, dt = \sum_{l=1}^{L}\frac{\Gamma_l}{L} sinc\left(\frac{\pi k}{L}\right) exp\left(\frac{i\pi k(2l-1)}{L}\right) \quad \text{(S5)}$$

Therefore, as indicated by Eq. (S5), the time-varying reflection coefficients $\Gamma(t)$ can be decomposed into the sum of a collection of complex exponential items, each of which is linked with a harmonic. To facilitate the understanding, we consider the magnitude and phase of ϵ_k as the synthetic reflection coefficient for the k^{th} harmonic wave. If $L = 1$, the spatiotemporal metasurfaces are degenerated into the basic spatial-modulated metasurfaces. This is consistent with Eq. (S5) because $\epsilon_k = 0$ with $k \neq 0$. If $L = 1$ and $k = 0$, Eq. (S5) is simplified to $\epsilon_0 = \Gamma_1$.

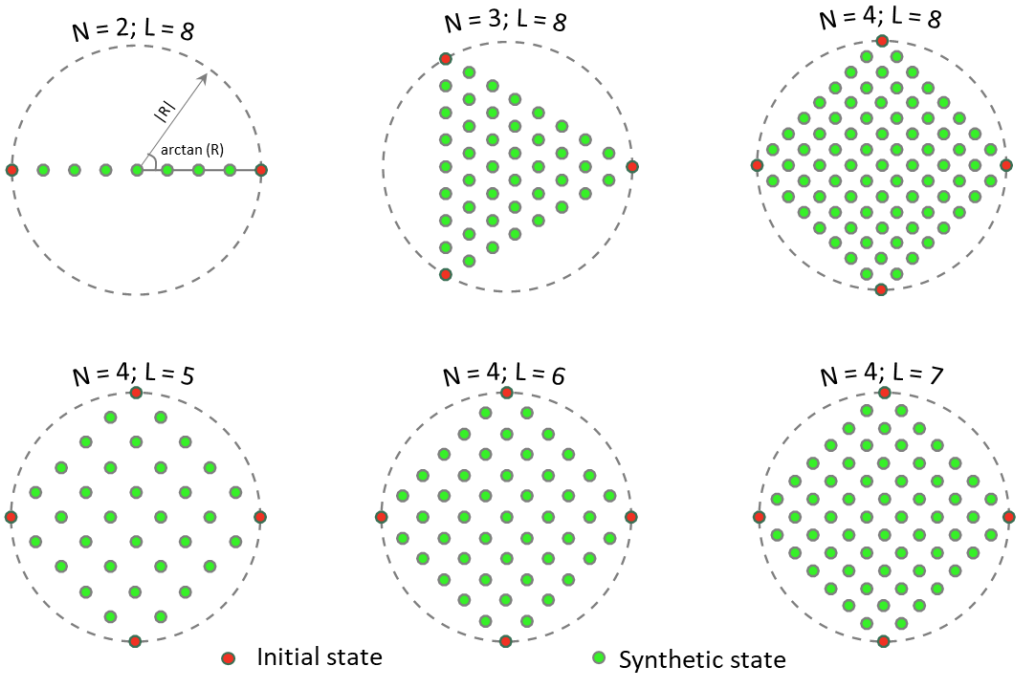

Figure S2 | Synthetic state with different initial state and different length of time-varying sequence. With the increase of the length of time-varying sequence, the number of synthetic states increase, gradually occupying the entire complex plane. The results in the figure are all at the main frequency. The distance between the synthetic state and the original point denotes the reflection amplitude, and the angle with the *x* axis denotes the reflection phase.

When L enlarges, the number of synthetic states increase. We note that a time-varying sequence can only induce a unique ϵ_k, but a ϵ_k can be induced by more-than-one time-varying sequences. We plot the synthetic state with different N and L, as shown in Fig. S2, where N is the number of initial states. When $N = 2$, we consider the amplitude of two initial states is unity, and the phase different is π. $L = 8$ will produce 2^8 time-varying series but with only seven synthetic states

(because many of them are overlapped) at the main frequency. Figure S2 exhibits a general trend that the number of synthetic states increase with the increase of N and L. In our work, we consider $N = 4$ and $L = 8$, and the synthetic states almost occupy the entire complex plane, which provide a high degree of freedom to freely manipulate electromagnetic waves.

For the same configuration of time-varying series, the synthetic states for different harmonics are different. Illustrated in Fig. S3 are the results with $N = 4$ and $L = 8$. It generally shows that the reflection amplitude is holistically compressed and becomes smaller for high-order harmonics. Notice that different harmonics are not completely decoupled.

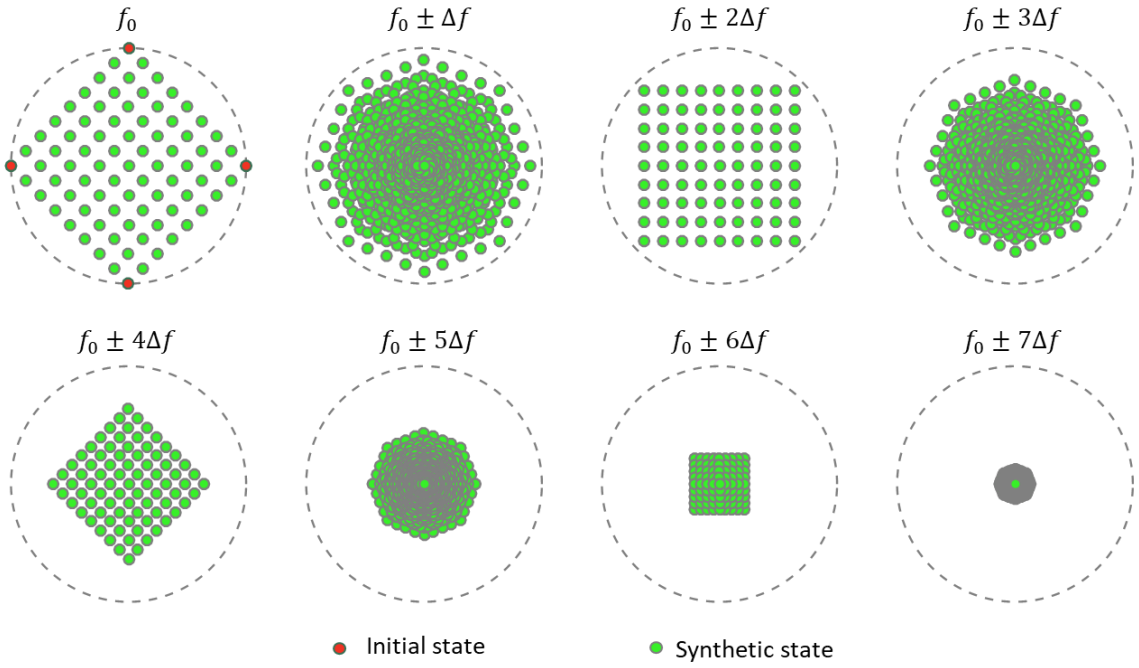

Figure S3 | Synthetic state in different harmonic waves. In this figure, $N = 4$ and $L = 8$. It is evident that the synthetic states gradually gather together. The amplitude of the synthetic state becomes smaller for high-order harmonic waves. f_0 is the frequency of incident wave, i.e., $m = 0$.

Supplementary Note 3: Comparison between spatial-only and spatiotemporal metasurfaces

To benchmark the superiority of spatiotemporal metasurfaces, we compare them with spatial-only metasurfaces for the same far-field customization task [S8]. For a given far-field, we optimize the profile of spatial-only and spatiotemporal metasurfaces using genetic algorithm (GA). The flowchart of GA is illustrated in Fig. S4a. The initial population is decoded into a group of metasurfaces, and the corresponding scattering performances are simulated and evaluated by minimizing a cost function. Then a genetic process (selective reproduction, crossing over, and mutation) is performed to update

the individuals until an optimal coding matrix is found. In Fig. S4b, we randomly generate three far-field patterns and mimic them with spatial and spatiotemporal metasurfaces (8 × 8), respectively. For spatial metasurfaces, each meta-atom has four discrete states (Supplementary Note 1) that can be chosen. For spatiotemporal metasurfaces, the states of each meta-atom can be freely picked from the synthetic states and initial states (Fig. S3). Evidently, the far-field pattern enabled by spatiotemporal metasurfaces is highly consistent with the target (in both shape and value), in sharp comparison with spatial metasurfaces. It suggests that, by using spatiotemporal metasurfaces, a more powerful ability in manipulating electromagnetic waves will be reached.

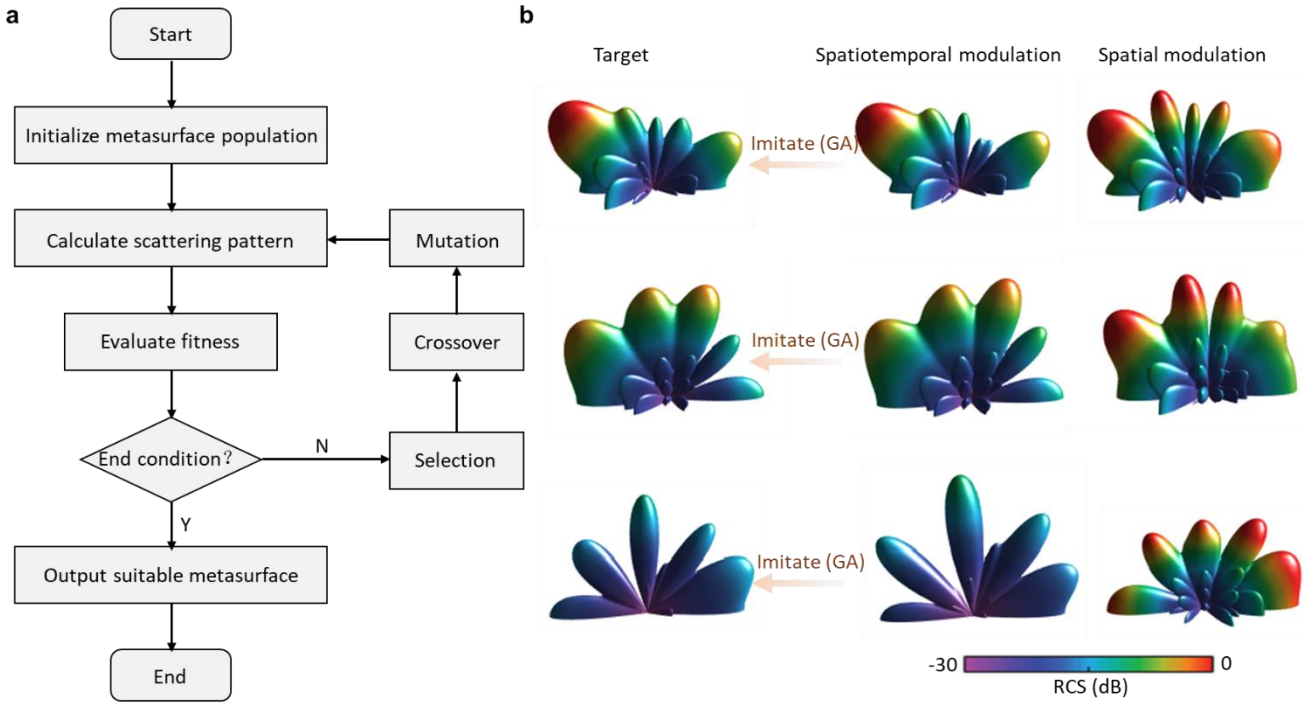

Figure S4 | Mimicking the far-field with spatial-only and spatiotemporal modulated metasurfaces. **a**, Flowchart of genetic algorithm. **b**, For the three randomly-given targets, we use spatiotemporal and spatial modulation to imitate them. We find that the spatiotemporal metasurfaces give a high fidelity not only in pattern shape but also in numerical value.

Supplementary Note 4: Broadband generalization of intelligent aeroamphibious cloak

The realization of broadband aeroamphibious cloak is a long-standing dream, albeit challenging. Although the cloaking evidence in our work has been demonstrated in a narrow working band, the proposed strategy of intelligent aeroamphibious cloak can also be easily generalized into broadband. To this end, foremost, we want to highlight that the broadband realization does not place a barrier

for intelligent algorithm, i.e., the generation-elimination network. We should add another frequency channel and then train the generation-elimination network in a similar manner. The difficulty is attributed to the metasurface physical performance and the modulation speed for spatiotemporal metasurfaces. In the following, we will specifically illustrate how to reach this goal.

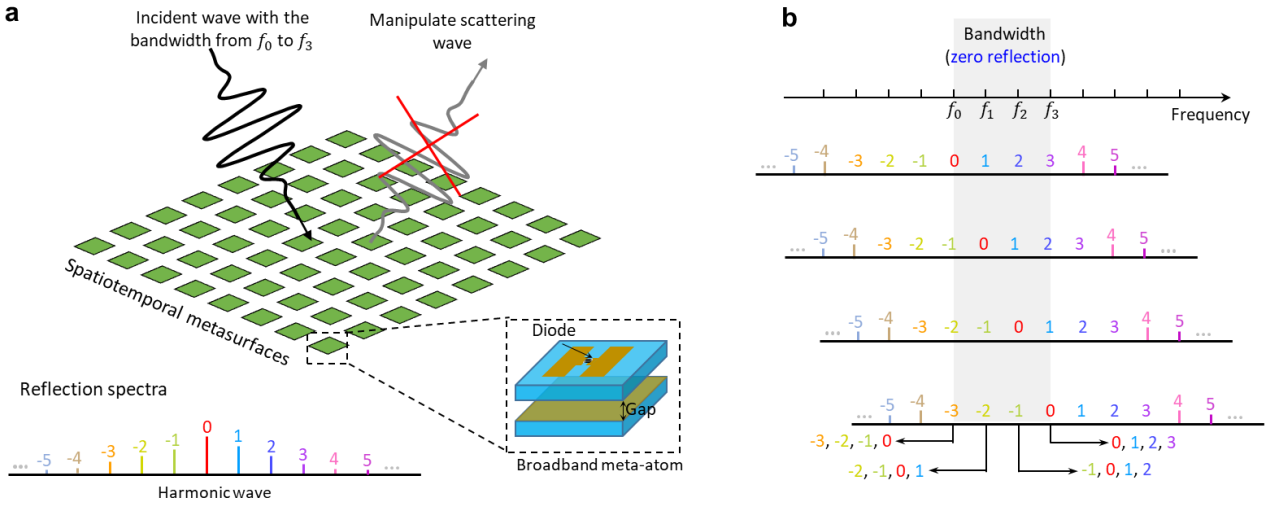

Figure S5 | Broadband realization of intelligent aeroamphibious cloak. a, The spatiotemporal metasurfaces suppress the scattering wave in broadband. Inset shows a broadband meta-atom. **b**, Principle of reaching zero reflection in broadband.

For example, under the illumination of incident wave with the bandwidth from f_0 to f_3, we aim to completely suppress the scattering wave (Fig. S5a). For the incident frequency component f_0, it will generate a series of harmonic waves that affect the scattering wave at other frequencies. Similarly, for any incident frequency component from f_0 to f_1, it will generate a series of harmonic waves. Thus, for the scattering wave at f_3, it is contributed by the zero-order harmonics induced by incident wave at f_1, first-order harmonics induced by incident wave at f_2, second-order harmonics induced by incident wave at f_1, third-order harmonics induced by incident wave at f_0, labelled as 0,1,2,3 (Fig. S5b). Similarly, for the scattering wave at f_0, it is contributed by -3,-2,-1,0. If we can exactly engine the reflection amplitude of these orders become zero, then the scattering field will be zero. Typically, the lower the order, the greater the reflection amplitude. However, we find it is possible to make these low-order harmonics become zero, which relies on the optimization of time-varying sequence and the physical properties of the meta-atom. Broadband and reconfigurable meta-atoms have been widely studied [S4]. For example, the double-layered meta-atom in the inset of Fig. S5a that

incorporates diode on a metallic resonator etched on dielectric substrate has great potential to work at broadband. After careful geometrical design, the reflection phase difference between on and off diode state is possible to maintain unchanged in broadband. This way, we just need to optimize the time-varying sequence to make all low-order harmonics become zero.

Supplementary Note 5: Generation-elimination network

The architecture of the generation-elimination network, constituted by an encoder, a latent space, a decoder, and a forward network, is schematically depicted in Fig. 3a with the detailed parameters listed in Fig. S6. As elucidated in the main text, the complete process can be divided into three steps: the pre-training of the forward network, the training of the whole generation-elimination network, and the inference phase.

The first step. Only the forward network composed of nine fully-connected layers is involved in this step. The detailed definition of the forward network is shown in the last three rows in Fig. S6.

	layer	name	op.	size-in	size-out	
	1	Input	-	22	-	Encoder
	2	Label	-	181	-	
Recognition module	3	Concat1	concatenation	22+181	203	
	4	FC1	fc relu	203	500	
	5	FC2	fc relu	500	500	
	6	FC3	fc relu	500	256	
	7	mu	fc linear	256	10	Latent space
	8	sigma	fc linear	256	10	
	9	Sampled	sampling	10	10	
	10	Label	-	181	-	Decoder
Reconstruction module	11	Concat2	concatenation	10+181	191	
	12	T_FC1	fc relu	191	500	
	13	T_FC2	fc relu	500	500	
	14	T_FC3	fc relu	500	200	
	15	Output	fc sigmoid	200	22	
	16	F_FC1	fc relu	22	512	Forward network
	17~23	F_FC2~F_FC8	fc relu	512	512	
	24	F_FC9	fc linear	512	181	

Figure S6 | Structure and parameters of the generation-elimination network. The recognition module combined with the *Input* and *Label* composes the encoder, while the reconstruction module combined with the *Label* composes the decoder. The latent space includes fully-connected operations for mean ("mu") and standard deviation units ("$sigma$") followed by the Gaussian sampling ("Sampled"). Both "FC" and "fc" refers to the fully-connected layer; "relu" refers to the ReLU activation function; "linear" refers to the linear activation function; "sigmoid" refers to the sigmoid activation function; "T_FC" is the abbreviation of transposed fully-connected layer and "F_FC" is the abbreviation of the fully-connected layer in the forward network.

The second step. The whole CVAE-based [S9] generation-elimination network participates in this core step. The parameters of the pre-trained forward network are fixed when training the entire network. As shown in Fig. S6, the recognition module is composed of three fully-connected layers, encoding the concatenation of the *Input* and *Label* into lower dimensions that are used to input into the latent space. The reconstruction module is composed of a concatenation operation and four transposed fully-connected layers, decoding the sampled latent variables into 20-dimensional design parameters.

The third step. There is no further training in the last inference step, and only the decoder and the forward network are involved. The rounding operation is carried out in this step. The sampled variables from the standard Gaussian distributions combined with the *Label* (i.e., the desired radar cross section, RCS) will be firstly decoded into countless candidates (i.e., 20-dimensional design parameters) [S10-S12]. Then, the design parameters will be rounded and transformed into 181-dimensional RCS value through the forward network. The best candidate is selected by finding the minimal RCS deviation with the *Label*. Here, we define the deviation as the Manhattan distance. Other metrics such as Euclidean distance and correlation distance could be adopted to screen out the corresponding best candidate or increase the loss on the main lobe if only focusing on that.

As elucidated in the third step and depicted in Fig. 3a, the generated candidates should be rounded before being sent into the forward network for further elimination. The 20-dimensional candidate is composed of 10 groups of [amplitude, phase]. The key point is that, because of the limited choices in the spatiotemporal modulation of metasurfaces, each of [amplitude, phase] group needs to be approximated to one of the 81 points (synthetic reflection states in Fig. 2c of the main text). Certainly, we should first convert [amplitude, phase] groups into the [real, imaginary] coordinate before the

approximation. As schematically illustrated in Fig. S7, the yellow point (one of the generated groups) will be approximated to the nearest green point (the limited choice) which has the minimum Euclidean distance with it. Besides, the RCS retrieved from the optimal candidate is also shown in Fig. S7 in either case. The little difference between the blue (the ground-truth) and orange (the prediction) curves in the after-rounding case verifies the practicability and effectiveness of our network.

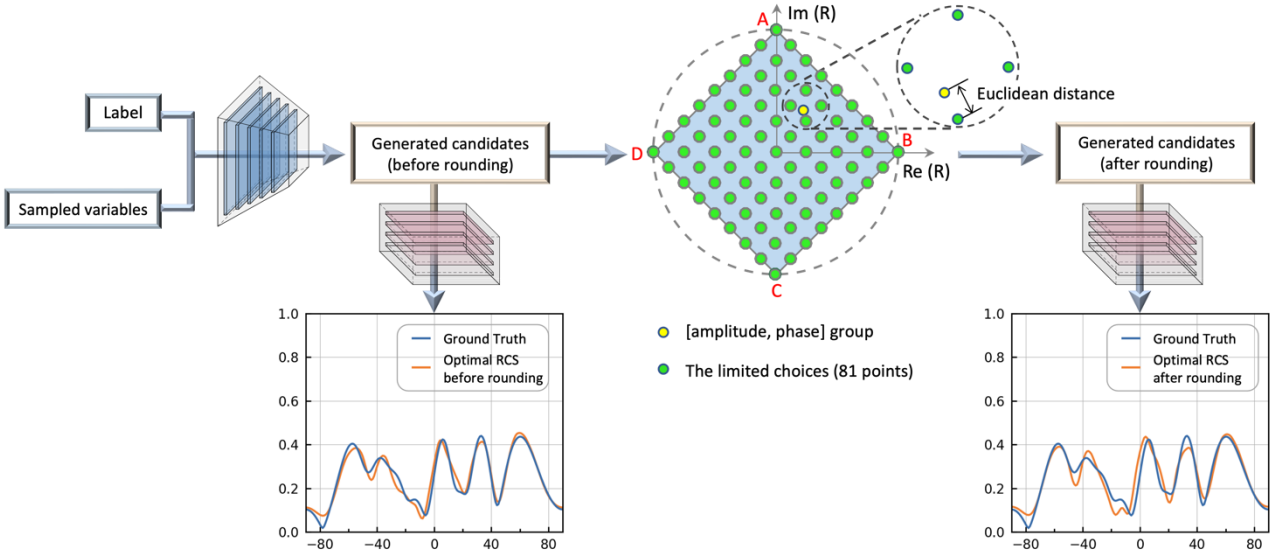

Figure S7 | Rounding operation in the inference phase. All generated candidates are composed of 10 [amplitude, phase] groups, each of which (the yellow point) will be approximated to the nearest choice (the green point) before being transformed into the RCS. The optimal after-rounding candidate is selected by finding the minimal RCS deviation between the blue curve and the orange curve in the right bottom plot.

Supplementary Note 6: Supervised learning loss and the accuracies

In the training of the forward network, it is automatically cut off at 211 epochs when the patience is set as 50 epochs as an early stopping regularization measure [S13]. As shown in Fig. S8, the validation loss reaches the minimal mean square error (MSE) of 1.28×10^{-4} at 161 epochs. Similarly, for the generation-elimination network with the patience set as 30 epochs, the training is automatically cut off at 239 epochs and the validation loss outputs the minimal value of 18.85 at 209 epochs. The objective loss function of the generation-elimination network is defined as:

$$\mathcal{L}(x, y; \theta, \varphi) = KL[q_\varphi(z|x, y)||\, p_\theta(z|y)] - \mathbb{E}_{q_\varphi(z|x,\, y)}[log\ p_\theta(x|z, y)] + \alpha(y - y')^2 \qquad \text{(S6)}$$

where the first term is KL divergence loss (evaluating the similarity between the approximate variational posterior and the prior probability), the second term is the reconstruction loss of x (calculated as the negative maximum likelihood) and the third term is the prediction loss of y (calculated as the MSE over the point). Notice that x is the 20-dimensional input, y is the 181-dimensional label variable and z is the latent variable. The deterministic RCS value of y' is predicted from the forward network based on the predictive distribution $P_p(y|x)$. The hyper-parameter α is set as 2,000.

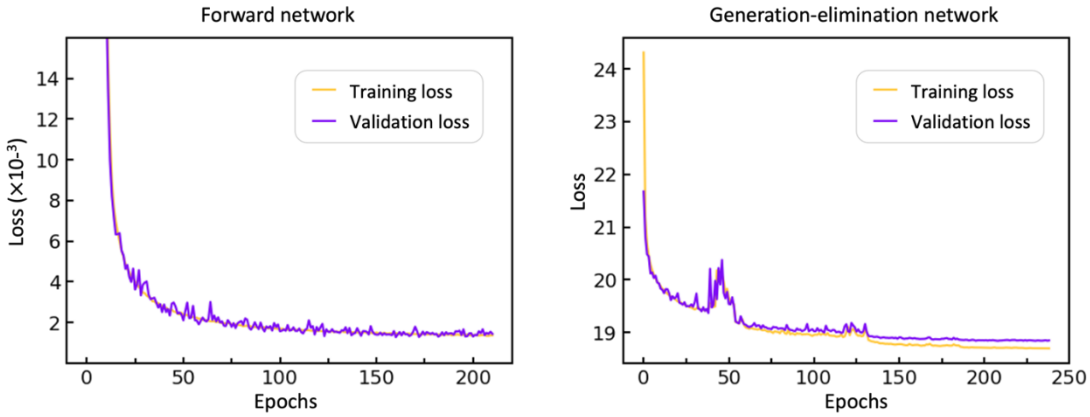

Figure S8 | The loss of forward network and generation-elimination network over epochs. a, The training and validation losses of the forward network. To avoid overfitting, the network is early-stopped at 161 epochs with the patience set as 50 epochs. **b,** The training and validation loss of the generation-elimination network. Similarly, the network is early-stopped at 209 epochs with the patience set as 30 epochs.

To quantity the performance of our network, we systematically define two criteria. (1) MSE, the mean square loss between the predicted RCS and ground truth RCS. (2) Accuracy $(1 - e_{ave}) \times 100\%$, where e_{ave} is defined as the average error between the predicted RCS and ground truth RCS, that is, $\frac{1}{n}\sum_{i=1}^{n} |y_i - y_i'|/y_i$, where $y_i(y_i')$ represents the ith data point of the ground truth (predicted) RCS, and n is the number of RCS points.

Figure S9 displays the summary statistics of two quantitative criteria when trained with the forward network and generation-elimination network, separately. Without reducing learning rate (reduceLr) measure [S14], the accuracy can be as high as 99.15% for the forward network. For the generation-

elimination network, the reduceLr measure does little help to the accuracy (from 97.67% to 97.68%), while increasing the dimension of the latent space improves the accuracy to some extent, from 96.86% of 2-dimension to 97.68% of 10-dimension. For comparison and reference, we also provide the accuracy of the predicted RCS that is selected without rounding operation on candidates, that is, the generated candidates are directly sent into the forward network for elimination. Further, we take other optimization measures such as batch normalization [S15] and dropout [S16], none of which can bring obvious improvement for the network performance.

		Forward network	Generation-elimination network			
Network configuration		no reduceLr	10-dim with reduceLr	10-dim no reduceLr	5-dim with reduceLr	2-dim with reduceLr
MSE		1.2807e-04	18.8473	19.2510	18.8578	18.8379
Accuracy	Before rounding	99.15%	98.27%	98.14%	98.16%	98.01%
	After rounding		97.68%	97.67%	97.55%	96.86%

Figure S9 | Two criteria to quantitatively evaluate the performance of networks upon different configurations. "reduceLr" refers to the reduce learning rate measure. "dim" is the abbreviation of dimension.

Supplementary Note 7: Gyroscope detector

In our system, an attitude sensor (HWT905-232) is applied to real-time recognize the drone's gesture so that the invisible drone can customize scattering wave in specific direction (Fig. S10a). The attitude sensor is a high-performance 3D motion attitude measurement system based on micro-electro-mechanical system (MEMS) technology, including three-axis gyroscope, three-axis accelerometer, three-axis electronic compass and other motion sensors. By integrating various high-performance sensors and attitude dynamics algorithm engines, the drone can be provided with a three-axis attitude angle (α, β, γ in Fig. S10b) with high accuracy (the measurement accuracy is 0.05°), high-dynamics and real-time compensation. The attitude sensor is connected with a six-in-one serial port conversion module to achieve USB-232 digital interface conversion and data serial input/output, which is driven by the CP2102 driver. To intuitively show the working effect of the attitude sensor in the experiment, we use the 3D attitude model in the built-in upper computer to illustrate the real-time attitude of the drone. Postures of three random moments are captured in Fig. S10c with

different attitude angles, which are further input into the generation-elimination network to inform the network of the invisibility requirement under different inclinations. The three curves in Fig. S10d show the changes of the acquired three-axis attitude angles, depicting the flight attitude of the drone.

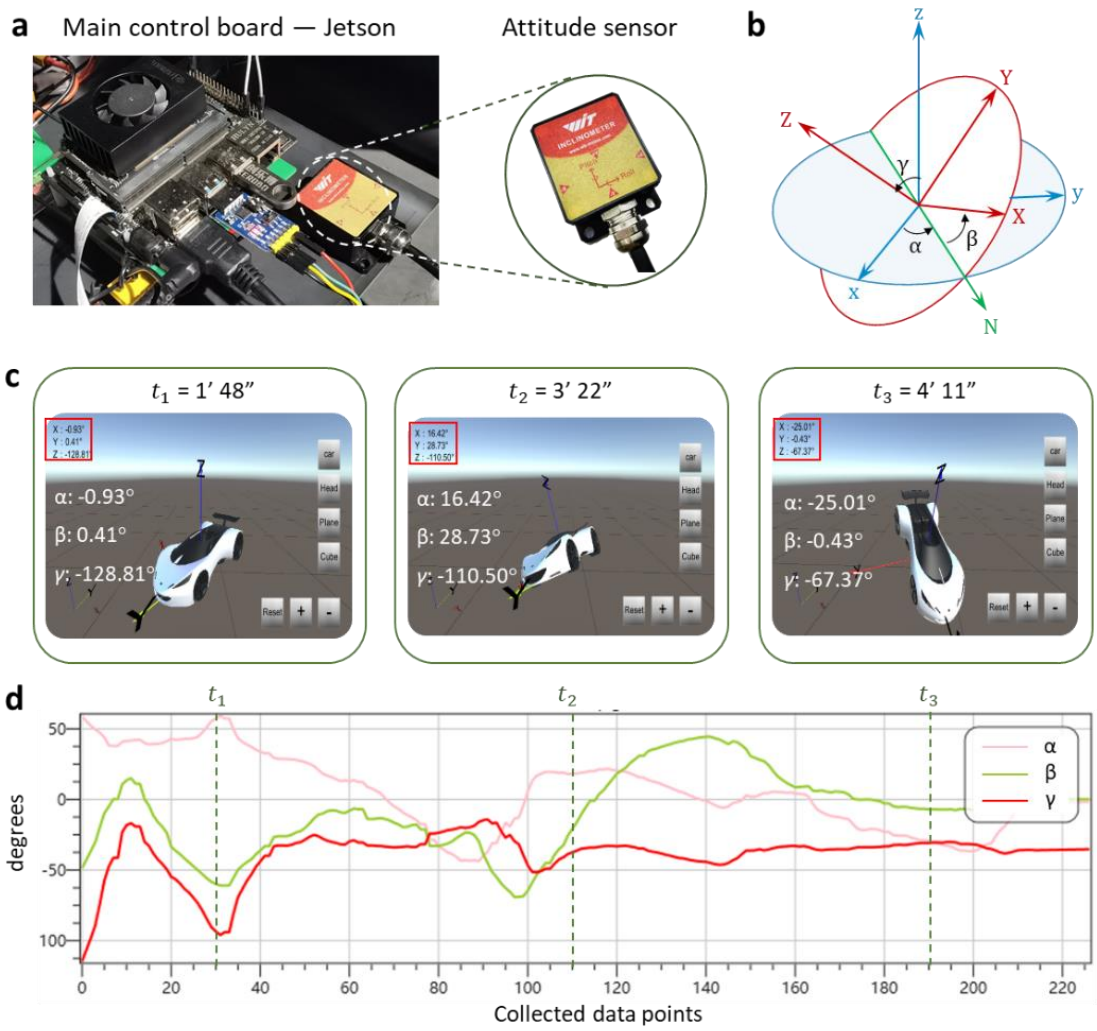

Figure S10 | Result of the attitude sensor. a, Diagram of the attitude sensor in the experiment setup, which is applied to recognize the drone's three-axis attitude angles in **b**. **c**, Posture of the three random moments captured by the 3D attitude model in the built-in upper computer. **d**, Acquired three-axis attitude angles.

Supplementary Note 8: Camera and environment discrimination network (EDN)

In the drone system, camera is another essential part of the perception module that is responsible for collecting and judging the ever-changing background. The type of camera we use is Raspberry Pi NoIR Camera v2 (Fig. S11a), which is specially designed for raspberry pie with IMX219 expansion board and is connected with CMOS Serial Interface (CSI) interface. To capture the real-time

environment, first, we use the CSI camera to achieve the transmission of rtsp stream and explore the dynamic acquisition of multiple streams. Then, we have the picture of the background environment in millisecond scale with the size of $1920 \times 1080 \times 3$, as shown in Fig. S11a, the background information of three random moments is recorded. Finally, the picture is resized into $32 \times 32 \times 3$ and further input into the EDN for environmental discrimination, which is constructed as a classification network. The structure of EDN is presented in Fig. S11b, mainly containing the feature extraction module (2 convolutional and 2 max-pooling layers) and the classification module (3 full connection layers). The output of EDN is one of four backgrounds coded as 0, 1, 2, 3. Adam optimizer is employed to update the parameters to complete the training of the model [17-18]. During training, the learning rate and batch size are set to 0.001 and 16, respectively. Forty environmental pictures are collected as training data, which can be divided into four types, including grass, cement, playground and water (Fig. S11c). The training result is shown in Fig. S11d, where the validation loss converges well with the training loss and the accuracy of EDN measured by the other 10 testing data is 100%.

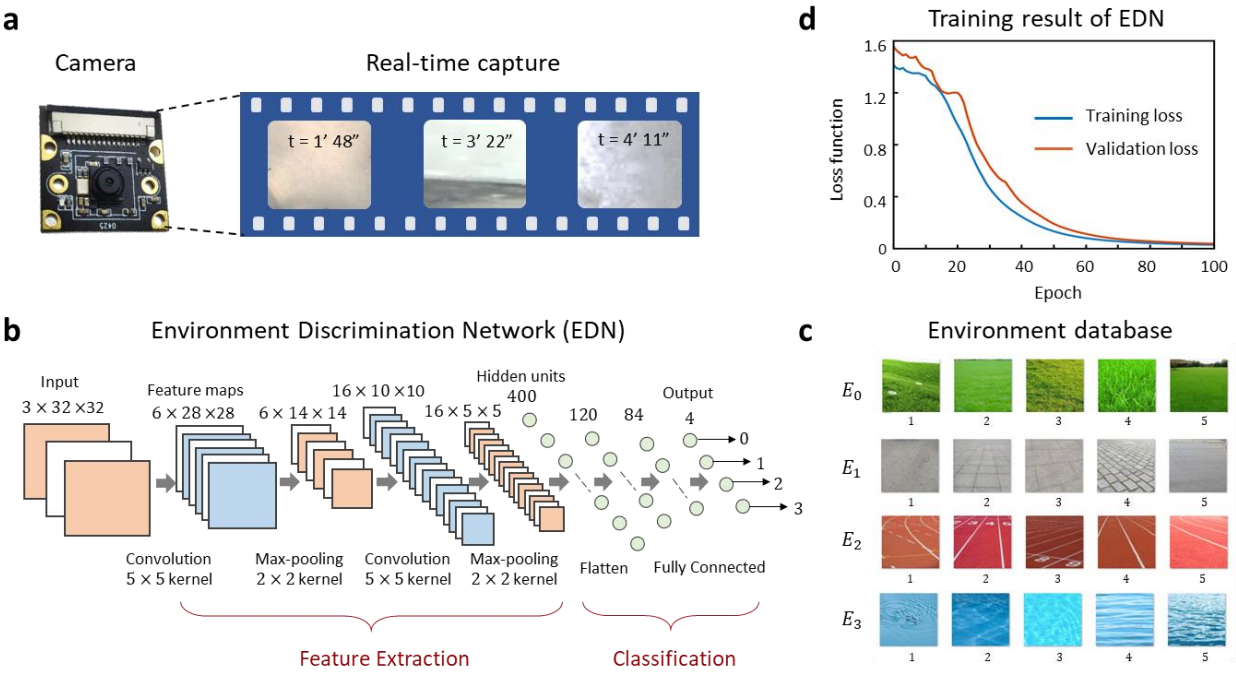

Figure S11 | Principle and function of the attitude sensor. a, Diagram of the CSI camera and the real-time environment capture at three moments. **b**, Detailed structure of EDN, which mainly contains feature extraction and classification module. **c**, Environment database of EDN, including four types: grass (E_0), cement (E_1), playground (E_2) and water (E_3). **d**, The training result of EDN.

Supplementary Note 9: Realization of an intelligent electromagnetic detector

A high-performance electromagnetic detector is essential for the invisible drone, which is used to perceive the information of incident wave, including direction of arrival (DOA) (the pitch angle θ, horizontal angle φ), frequency, and polarization in the microwave band. The schematic view of the proposed intelligent electromagnetic detector and its operation principle are displayed in Fig. S12. The detector mainly consists of a four-port wide-band antenna array, an RF processor (AD9361), and an algorithm processing platform (ZYNQ). The antenna used here is a wide-band coplanar waveguide (CPW) antenna, whose working frequency is set from 2 GHz to 4 GHz. Figure S13 shows the structure and the S_{11} parameters during this frequency band (< -10 dB). Four antenna elements are printed on an octagon substrate using printed circuit board (PCB) technology. The dielectric substrate is made of F4B material (ε = 4.4) and the thickness is 1 mm. Antenna elements are placed at a 90-degree rotation interval to constitute an omnidirectional antenna array and curtail the mutual coupling.

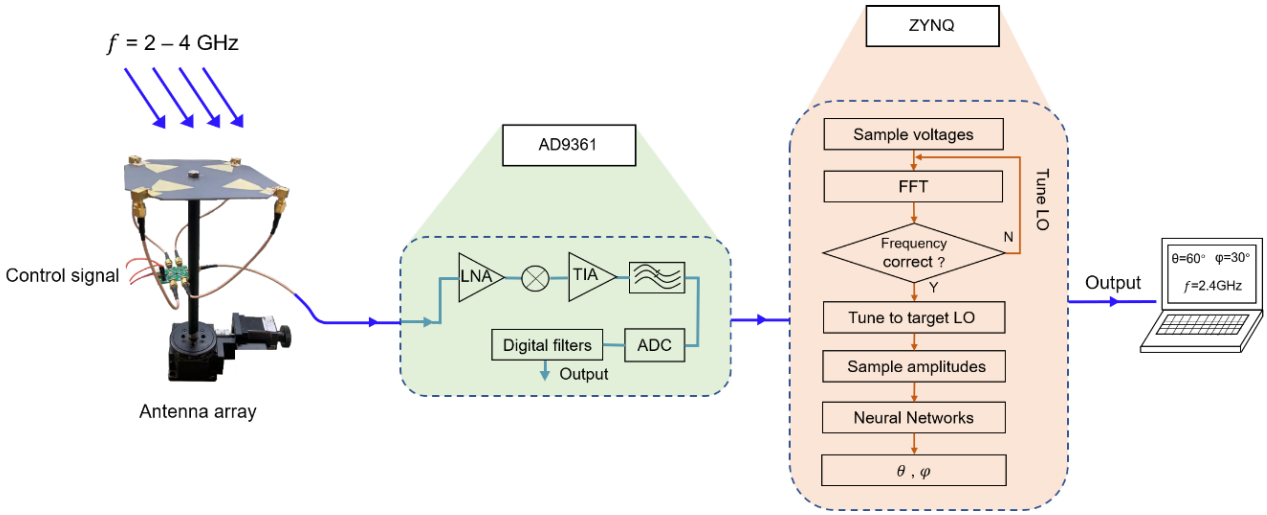

Figure S12 | Working flowchart of the intelligent electromagnetic detector. For an unknown incident wave, the antenna array receives electromagnetic signal, which is then processed by AD9361 (a software defined radio (SDR) platform), containing low-noise amplifier (LNA), mixer, analog/digital filter, analog-to-digital converter (ADC), etc. After the ADC sampling, the voltage sequence is transferred to the ZYNQ platform, where the machine learning model is intergrated. Finally, the information of the incoming wave is directly displayed on an user interface in a millisecond timescale.

From the flowchart in Fig. S12, an electromagnetic wave impinging on the antenna array will induce the voltage on each antenna element. For different incident wave, the induced voltages are different, making it possible to inversely determine the incident wave by the induced voltages. The four-port

voltages from the antenna array are transmitted into radio frequency (RF) processor. To mitigate the complexity and the cost of the electrical system, RF switch is used to serially read four channels from the antenna array. ZYNQ controls the SP4T RF switcher (HMC7992) to sample the signal from four different ports and make the AD9361 to receive it. The RF processor is used for amplification, analog filtering, and down converting of the input signal and then passes it to the ADC for sampling.

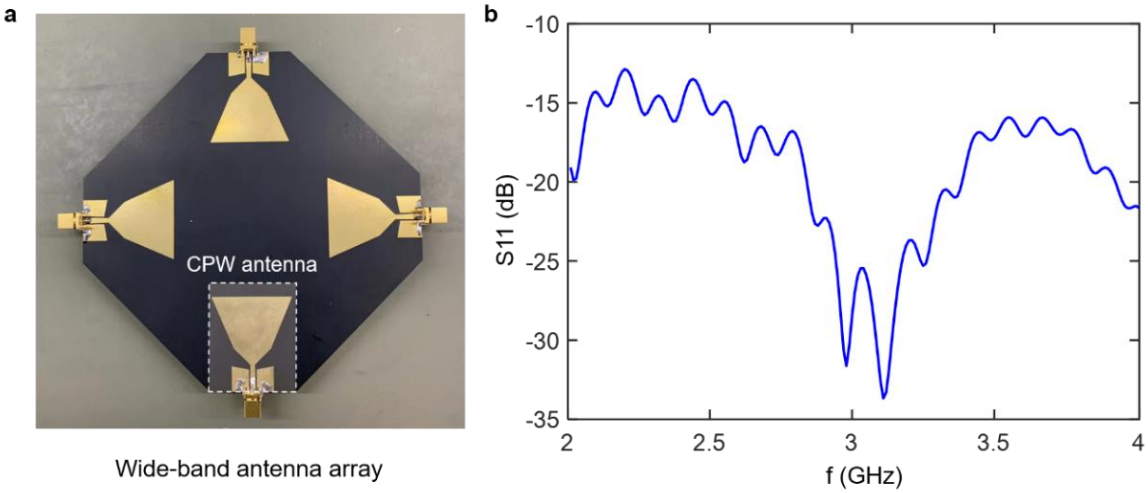

Figure S13 | Details of wideband antenna array. **a**, Top view of the antenna array. The antenna array is composed of four CPW antenna. **b**, The S_{11} parameter of the CPW antenna element.

After the processing by AD9361, the first step in ZYNQ is to obtain the frequency component of the incoming wave. This can be readily retrieved by frequency sweep and Fourier transform. Then, the voltage vector will be fed to the neural network for the DOA and polarization determination. Here, we deploy the generalized regression neural network (GRNN) to inversely deduce the information of one incident source [S19,S20]. As a member of radial basis neural networks, GRNN has a strong nonlinear approximation ability, making it suitable for this function fitting like task. In one of our previous works, we have verified the feasibility of the GRNN method. Here, we just skip the detailed description of this method.

We note that the entire detection takes about 25 ms, among which 18 ms is for frequency sweeping over broadband, 2 ms is for the neural network calculation, and 5 ms is consumed by other data-processing algorithms, such as fast Fourier transform and median filter. Compared with conventional spatial spectrum methods, one advantage of this intelligent electromagnetic detector is that, we only

consider amplitude information so that phase disturbances introduced by the switches and other miscellaneous components will not greatly affect the experimental accuracy.

Supplementary Note 10: Control system of the spatiotemporal metasurfaces

In total, four metasurface boards are covered over the drone to render it invisible, including top board (10 x 10 unit cells), bottom board (10 x 11 unit cells) and two side boards (8 x 11 unit cells). Each unit cell on the same column is connected with the same signal lines and thus shares the same state. There are two signal lines to control the two PIN diodes inside the unit cell, allowing four states (on, on), (on, off), (off, on) and (off, off), as shown in Fig. S14.

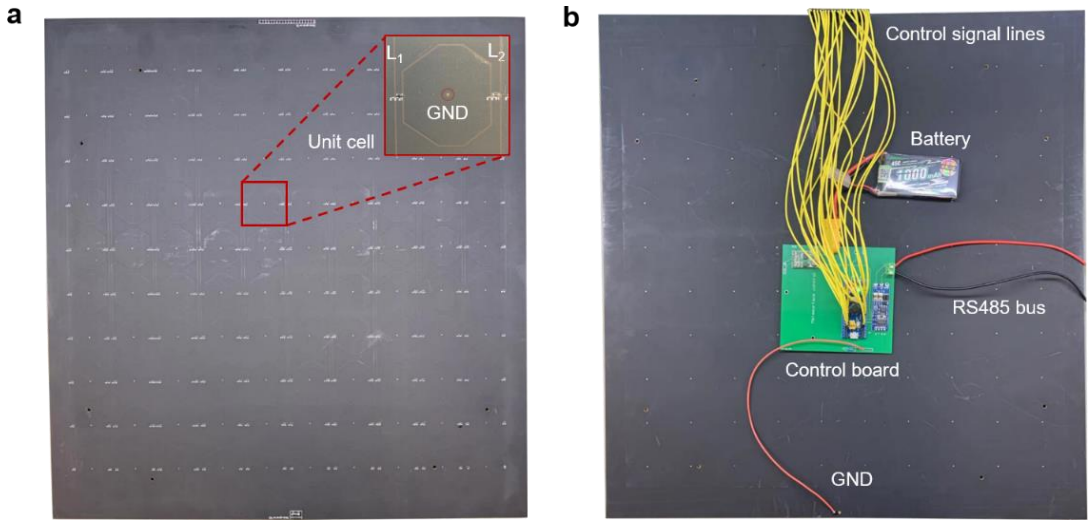

Figure S14 | Circuit design of the metasurface board. **a**, The top view of the metasurface board. Each column of the metasurfaces own two signal lines (L_1 and L_2), allowing four discrete reflection states. **b**, The bottom view of the metasurface board. The control signal lines are connected with STM32 chip on the control board. The control board first receives the signal sequence and then outputs the signal by the I/O on the STM32 chip.

The PIN diode is controlled by the microcontroller unit (MCU). We adopt the STM32 series chip to output the periodic time-varying voltage sequence in microsecond. To ensure a fast switch speed and to improve the scalability of the system, four MCUs are connected via one bus (RS485). Benefitting on the differential voltage transmission, the RS485 bus owns strong anti-interference ability and can realize a long-distance transmission. Additionally, it can mount up to 32 MCUs on one bus, which is suitable for the large-scale system. The bus is controlled from the deep learning hardware platform—Jetson Xavier. After generating the spatiotemporal sequence by the pre-trained generation-

elimination network, Jetson sends out the signal using universal asynchronous receiver/transmitter (UART) port, which is immediately transferred to RS485 communication protocol as the input of bus.

Supplementary Note 11: Working flowchart of intelligent invisible drone

The intelligent visible drone needs multiple sensors to recognize the surrounding environment and incoming wave. These sensors mainly include a gyroscope, a camera, and an electromagnetic detector. The gyroscope (HWT905-232) is used to real-time judge the attitude, acceleration speed, and angular velocity of the drone itself. The camera is used to dynamically recognize the surrounding environment, such as grassland, water area, desert, and cement floor.

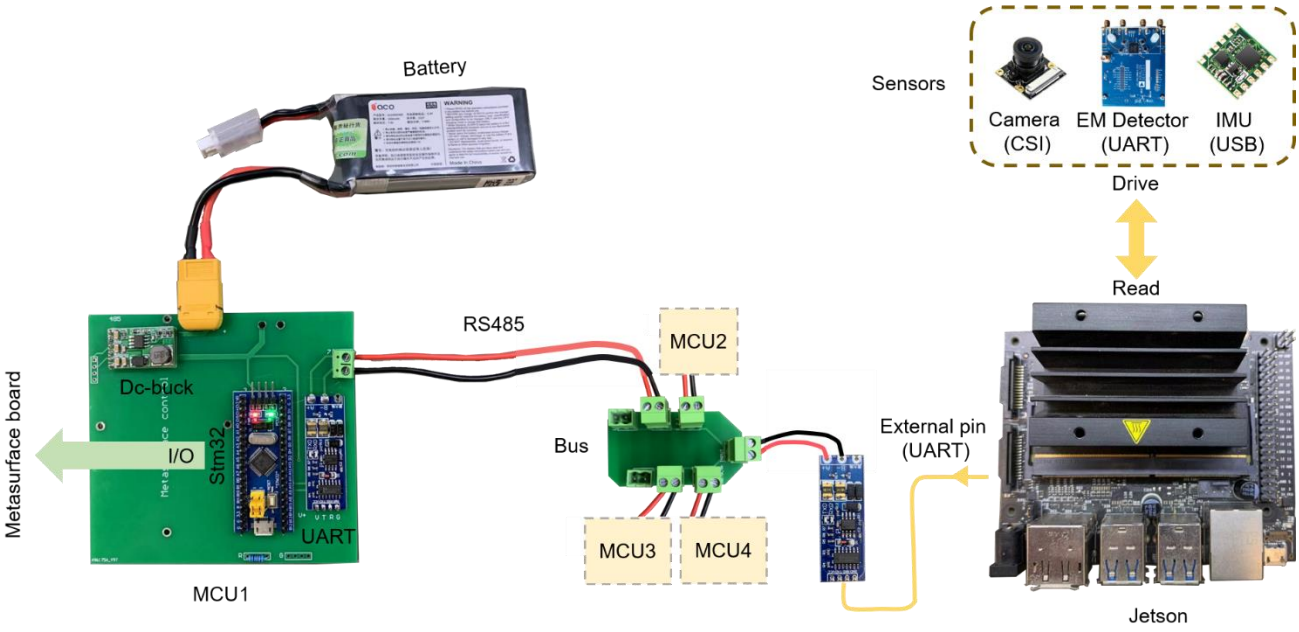

Figure S15 | Control system of the intelligent invisible drone. From the right to the left, Jetson first reads the detection information from the sensors and then sends it to the pre-trained neural network. The neural network outputs the spatiotemporal voltage sequence by the UART port. The signal is converted to the RS485 communication protocol and distributed to each metasurface sub-system (four in total) using RS485 bus. Finally, the signal will be received by the UART port on the STM32, which is applied to the metasurface board using input/output (I/O). IMU, inertial measurement unit. UART, universal asynchronous receiver/transmitter.

As the core control board of the invisibility system, the whole operation process of Jetson is depicted in Fig. S15, which can be divided into perception, decision, and execution modules. For the perception module, first, we open the CSI camera and realize the transmission of real-time streaming protocol

(RTSP) stream to dynamically acquire multiple streams for exploration. Then, the Jetson captures the ever-changing background and save it as a picture, which is further input into the environment discrimination network (EDN) for environmental discrimination. Third, Jetson reads the dynamic 3D attitude of the drone from the gyroscope through USB interface. Combined with the detected incident wave's direction through the EM detector, the information of all these modules is integrated into the real-time invisibility demand and further input into the pre-trained generation-elimination network. After seeking out the best output candidate, the optimal time-varying sequences for spatiotemporal metasurfaces are then converted to voltage signals for all metasurface boards. To the execution module, the time-varying sequence will be output to the bus and received by the STM32 on each board. The MCU then outputs the voltage signal through I/O interface to dynamically transform the states of the diodes on metasurfaces following the spatiotemporal signal.

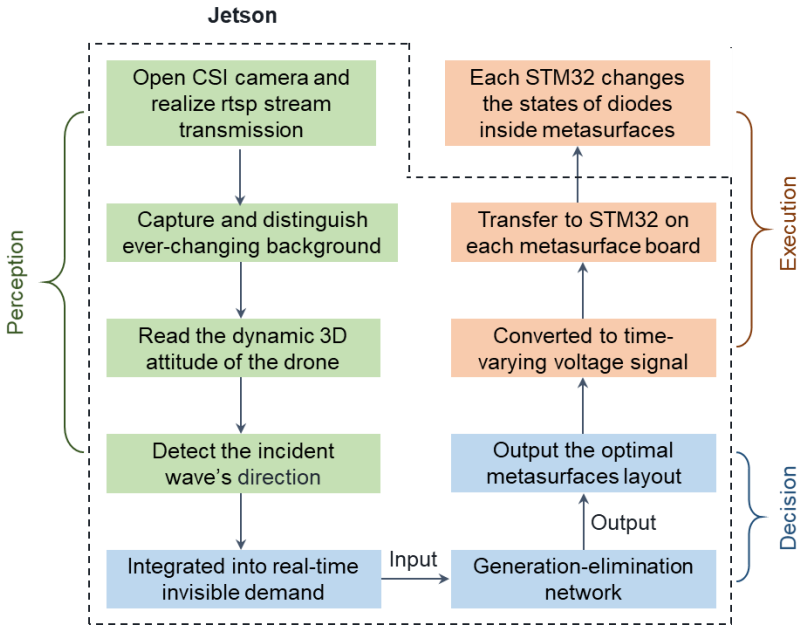

Figure S16 | Flow chart of the operation process of Jetson. As the core control board of the invisibility system, the whole operation process of Jetson can be divided into perception, decision, and execution modules, for self-adapting to the ever-changing background and detection manner.

Supplementary Note 12: Experiment setup of invisible drone against amphibious background

The experiment of the intelligent invisible drone on the land/sea is carried out in an anechoic chamber, mainly including a transmitting horn antenna, receiving antennas (probes), an intelligent EM detector and a Jetson development board (NVIDIA Jetson Xavier NX) [S21-S24], as shown in Figs.

S17a and S17b. The transmitting antenna is mounted on a lever that is bound to the base of the large triangle bracket, while 8 detectors are fixed on the small tripods. The transmitting antenna and all probes are arranged on the x plane, where the transmitting antenna is suspended on the top of the drone at a distance of 2 m. Assuming the top of the drone is at $z = 0$ plane and the center is the coordinate origin, we further adjust the height of 8 detectors to detect 8 points in the far-field section of $y = 0$ plane (Fig. S17c). Through measurement and calculation, 8 detectors are located at angles of $\delta =$ -68°, -50°, -38°, -24°, 24°, 41°, 50°, 65°. To mimic the on-site environment of grassland/sand-land/sea, we put acrylic containers on the absorbing materials, which are used to contain turfs/soil/water, as shown in Fig. S17d-S17f. To build a set of intelligent invisibility system, an intelligent EM detector composed of a four-port antenna array is used for the simultaneous attainment of frequency, direction-of-arrival and polarization. In conjunction with the camera and attitude sensor, we can output the time-varying sequence of spatiotemporal metasurfaces in real-time for an arbitrary invisibility requirement.

The brief process of the invisibility experiment is as follows. First, we measure the far-field of all backgrounds at the desired frequency. Second, we turn on the Jetson to automatically detect the attitude of the drone and identify the background environment. Third, the system processes the in-situ invisibility RCS requirement according to the measured background and attitude information. Then, we feed the preprocessed RCS and the perceived incident wave information, into the pre-trained generation-elimination network, which automatically outputs the corresponding time-varying sequence of spatiotemporal metasurfaces. The RCS of the cloaked drone is further measured and compared with the background RCS, while that of the bare drone is also tested for comparison. The operation procedure of the illusive experiment is similar to the above [S25-S27]. First, we measure the RCS of an illusive object at the desired frequency. Combined with the perception module, we input the illusive RCS into the decision module, which outputs the corresponding metasurface pattern. Finally, the RCS of spatiotemporal metasurfaces is measured. In Fig. 5 of the main text, we can observe that the detected value of the background and cloaked drone is consistent with each other, regardless of the attitude change of the drone itself and the change of external environment.

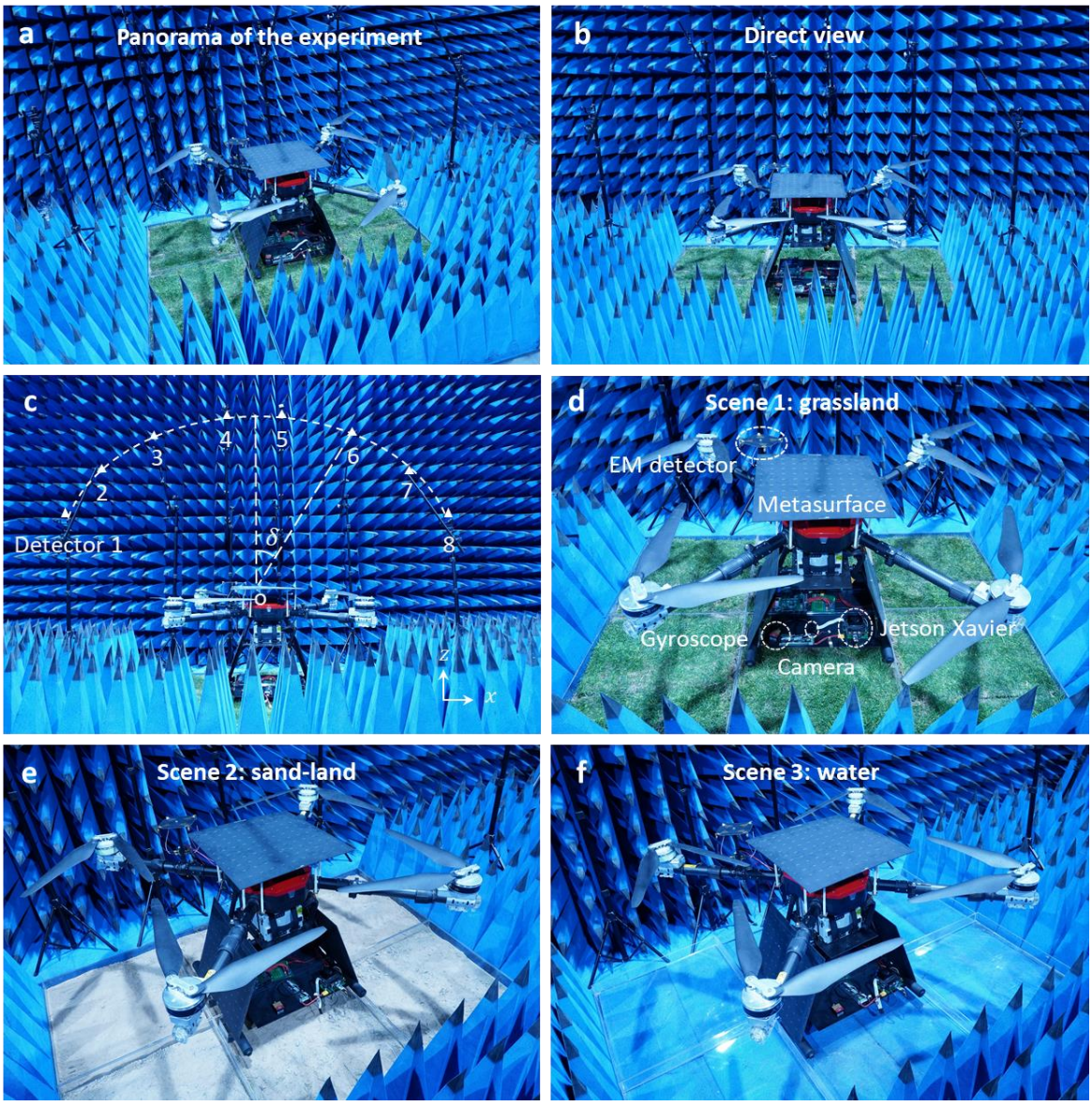

Figure S17 | Experimental setup of the invisible drone. (**a**) Panorama and (**b**) direct view of the experiment setup. **c**, 8 detectors are fixed on the small tripods at angles of δ = -68°, -50°, -38°, -24°, 24°, 41°, 50°, 65°, while the distance between probes and the origin o is about 1.5 m. To mimic the environments of (**d**) grassland, (**e**) sand-land, and (**f**) sea, we put acrylic containers on the absorbing materials, which are used to hold turfs/soil/water.

Supplementary References

[S1] Zhao, J. et al. A tunable metamaterial absorber using varactor diodes. *New J. Phys.* **15**, 043049 (2013).

[S2] Yang, H. et al. A programmable metasurface with dynamic polarization, scattering and focusing control. *Sci. Rep.* **6**, 35692 (2016).

[S3] Chen, K. et al. A reconfigurable active Huygens' metalens. *Adv. Mater.* **29**, 1606422 (2017).

[S4] Zhang, N. et al. Spatiotemporal metasurface to control electromagnetic wave scattering. *Phys.*

Rev. Appl. **17**, 054001 (2022).
[S5] Zhang, L. et al. Dynamically realizing arbitrary multi-bit programmable phases using a 2-bit time-domain coding metasurface. *IEEE Trans. Antenn. Propag.* **68**, 2984-2992 (2019).
[S6] Taravati, S., & Eleftheriades, G. V. Microwave space-time-modulated metasurfaces. *ACS Photon.* **9**, 305-318 (2022).
[S7] Hadad, Y., Sounas, D. L. & Alù, A. Space-time gradient metasurfaces. *Phys. Rev. B* **92**, 100304(R) (2015).
[S8] Balanis, C. Antenna theory (Wiley, 2016).
[S9] Sohn, K., Lee, H. & Yan, X. Learning structured output representation using deep conditional generative models. *Adv. Neural Inf. Process. Syst.* **28**, 3483-3491 (2015).
[S10] Kingma, D. P., Rezende, D. J., Mohamed, S. & Welling, M. Semi-supervised learning with deep generative models. *Adv. Neural Inf. Process. Syst.* **27** (2014).
[S11] Kingma, D. P. & Welling, M. Auto-encoding variational bayes. In *Proc. 2nd Int. Conf. Learn. Represent.* (ICLR, 2014).
[S12] Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *Int. Conf. Mach. Learn.* **32**, 1278-1286 (2014).
[S13] Yao, Y., Rosasco, L. & Caponnetto, A. On early stopping in gradient descent learning. *Constr. Approx.* **26**, 289-315 (2007).
[S14] Smith, L. N. Cyclical learning rates for training neural networks. *IEEE Winter Conf. Appl. Comput. Vis.*, 464-472 (WACV, 2017).
[S15] Ioffe, S. & Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *Int. Conf. Mach. Learn.* **37**, 448-456 (2015).
[S16] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929-1958 (2014).
[S17] Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **61**, 85-117 (2015).
[S18] Genty, G. et al. Machine learning and applications in ultrafast photonics. *Nat. Photon.* **15**, 91-101 (2021).
[S19] Specht, D. F. A general regression neural network. *IEEE Trans. Neural Netw.* **2**, 568-576 (1991).
[S20] Huang, M. et al. Machine-learning-enabled metasurface for direction of arrival estimation. *Nanophotonics* **11**, 2001-2010 (2022).
[S21] Qian, C. et al. Experimental observation of superscattering. *Phys. Rev. Lett.* **122**, 063901 (2019).
[S22] Wu, N., Jia, Y., Qian, C. & Chen, H. Pushing the limits of metasurface cloak using global inverse design. *Adv. Opt. Mater.* **2202130**, 1-8 (2023).
[S23] Wang, Z. et al. Reconfigurable matrix multiplier with on-site reinforcement learning. *Opt. Lett.* **47**, 5897-5900 (2022).
[S24] Zhu, R. et al. Phase-to-pattern inverse design paradigm for fast realization of functional metasurfaces via transfer learning. *Nat. Commun.* **12**, 2974 (2021).
[S25] Wang, X. & Caloz, C. Spread-spectrum selective camouflaging based on time-modulated metasurface. *IEEE Trans. Antenn. Propag.* **69**, 286-295 (2021).
[S26] Tennant, A. & Chambers, B. Time-switched array analysis of phase-switched screens. *IEEE Trans. Antennas Propag.* **57**, 808-812 (2009).
[S27] Fan, S. W. et al. Reconfigurable curved metasurface for acoustic cloaking and illusion. *Phys. Rev. B* **101**, 024104 (2020).