Appendices

Section I: The principles of spin-multiplexing metasurface

We all know that polarized light incident on the metasurface can usually be decomposed into the corresponding LCP and RCP spin eigenstates. In order to achieve different phase modulations, the spin-multiplexing metasurface needs to provide independent phase profiles $\varphi_1(x,y)$ and $\varphi_2(x,y)$ for LCP and RCP respectively. That is, for LCP incident light, the metasurface implements transformation: $\left|\mathrm{LCP}\right\rangle \to \exp(i\varphi_1)\left|\mathrm{RCP}\right\rangle$, the output beam is RCP polarized light with opposite handedness. Similarly, corresponding to RCP incident light, $\left|\mathrm{RCP}\right\rangle \to \exp(i\varphi_2)\left|\mathrm{LCP}\right\rangle$. Therefore, the metasurface can be described by the Jones matrix $J(x,y)$ satisfying $J(x,y)\left|\mathrm{LCP}\right\rangle = \exp(i\varphi_1)\left|\mathrm{RCP}\right\rangle$ and $J(x,y)\left|\mathrm{RCP}\right\rangle = \exp(i\varphi_2)\left|\mathrm{LCP}\right\rangle$. Therefore, the required Jones matrix takes the form:

$$J(x,y) = \begin{bmatrix} \dfrac{\exp\left[i\varphi_1(x,y)\right] + \exp\left[i\varphi_2(x,y)\right]}{2} & \dfrac{i\exp\left[i\varphi_2(x,y)\right] - i\exp\left[i\varphi_1(x,y)\right]}{2} \\ \dfrac{i\exp\left[i\varphi_2(x,y)\right] - i\exp\left[i\varphi_1(x,y)\right]}{2} & \dfrac{-\exp\left[i\varphi_1(x,y)\right] - \exp\left[i\varphi_2(x,y)\right]}{2} \end{bmatrix}. \tag{1}$$

According to the symmetry and unitary conditions, the $J(x,y)$ can be written in the standard form $J(x,y) = R\Lambda R^{-1}$, where R is a real unitary matrix and Λ is a diagonal matrix.

For birefringent nanopillars, the diagonal matrix Λ determines their phase shifts $\varphi_x(x,y)$ and $\varphi_y(x,y)$ along the two perpendicular symmetry axes, while R determines the rotation angle θ of their fast axis relative to the x-axis. (φ_x,φ_y) and θ respectively corresponds to the propagation phase and geometric phase imposed on the transmitted light. For given spin multiplexing phases $\varphi_1(x,y)$ and $\varphi_2(x,y)$, the required phase and rotation angle are calculated as follows:

$$\begin{cases} \varphi_x(x,y) = \frac{1}{2}\left(\phi_1(x,y) + \phi_2(x,y)\right) \\ \varphi_y(x,y) = \frac{1}{2}\left(\phi_1(x,y) - \phi_2(x,y)\right) + \pi \\ \theta = \frac{1}{4}\left(\phi_1(x,y) - \phi_2(x,y)\right) \end{cases} . \tag{2}$$

Therefore, in order to realize the Jones matrix $J(x,y)$, a series of subwavelength nanopillar should be designed to cover the required phases φ_x and φ_y, and meet the orientation angle θ at the point (x, y). The propagation phases φ_x and φ_y imparted by a single nanocolumn as a half-wave plate can be determined by the in-plane dimensions along two perpendicular symmetry axes, and the geometric phase is controlled by the steering angle θ of the nanocolumn.

Section II: Network training methods of diffractive neural network

The metasurface is constructed employing a training methodology akin to conventional electronic neural networks. The Huygens-Fresnel principle posits that the metasurface units within each layer function as sources for the subsequent layer's structure, resembling the process of neuron connections in electronic neural networks. According to the Rayleigh-Sommerfeld diffraction integral and the Jones matrix Fourier optics principle, the light field of the l layer of the all-optical neural network can be expressed as:

$$U\left(r^{l+1}\right) = \iint_{-\infty}^{+\infty} U\left(r^{l}\right) \cdot J\left(r^{l}\right) \cdot h\left(r^{l+1} - r^{l}\right) dxdy \tag{3}$$

where $U(r^l)$ represents the light field of the l layer, $J(r^l)$ represents the interaction matrix of the metasurface, and $h(r^{l+1}-r^l)$ is the factor of the Rayleigh-Sommerfeld diffraction integral, r is the spatial position of the metasurface. Then a forward propagation diffraction neural network model can be constructed. We define the loss function as:

$$F\left(\varphi_i^l\right)=\frac{1}{K}\sum_k\left(s_o^N-g_o^N\right)^2,\ \text{s.t. } 0\leq\phi_i^l\leq 2\pi \tag{4}$$

where s_o^N is the target field and g_o^N is the real field distribution. By employing the loss function and stochastic gradient descent algorithm, the network can be trained to iteratively update the action matrix of the hypersurface in order to minimize the loss function.

For the diffractive neural network, the training and testing process is completed by Python v3.11.0 and Pytorch v2.1.0. The DNN training framework runs on a GeForce RTX 4070Ti graphics processing unit (GPU, Nvidia Inc.) and an Intel® Core™ i7-12700 CPU @2.10 GHz central processing unit (CPU, Intel Inc.). The optical response of each nanopillar, including transmittance and phase, was calculated numerically using the finite difference time domain (FDTD) algorithm. In-plane periodic boundary conditions and vertical perfectly matched layers are applied.

Section III: The Detailed results about patterns detection of high-order VSBs

We consider left-handed circular polarization and right-handed circular polarization, and two sets of single-channel simulation results are shown in Fig. S1-S2. The metasurface is illuminated by a structured beam consisting of 36 Laguerre Gaussians (LG) beams, which have azimuthal indexes *l* from -4 to +4 and a radial indexes *p* from 0 to 3. At the output, two channels are designed with a basic Gaussian pattern, appearing on the left and right sides of the LCP and RCP screens respectively. On the left side of the screen, 36 output channels are generated for the LCP component, carrying additional azimuthal topological charges l = -4, -3…+3, +4 from bottom to top, and carrying radial topological charges *p* = 0, 1, 2, 3 from left to right. Each channel forms a basic Gaussian highlight on the screen. The output channels of the RCP components are symmetrical relative to the output channels of the LCP components.

Furthermore, in order to quantitatively evaluate the performance of pattern detection, the energy ratio of 36 channels on the output plane was calculated, as shown in Fig. S3-S4. The energy ratio was normalized by dividing the light intensity of each channel by the total light intensity of the 36 channels. It can be seen from the results that the average energy percentage of the output channel exceeds 80% and the crosstalk is close to 10.5dB, as shown in Fig. S5. The research results show that the device has the characteristics of high accuracy and small crosstalk.

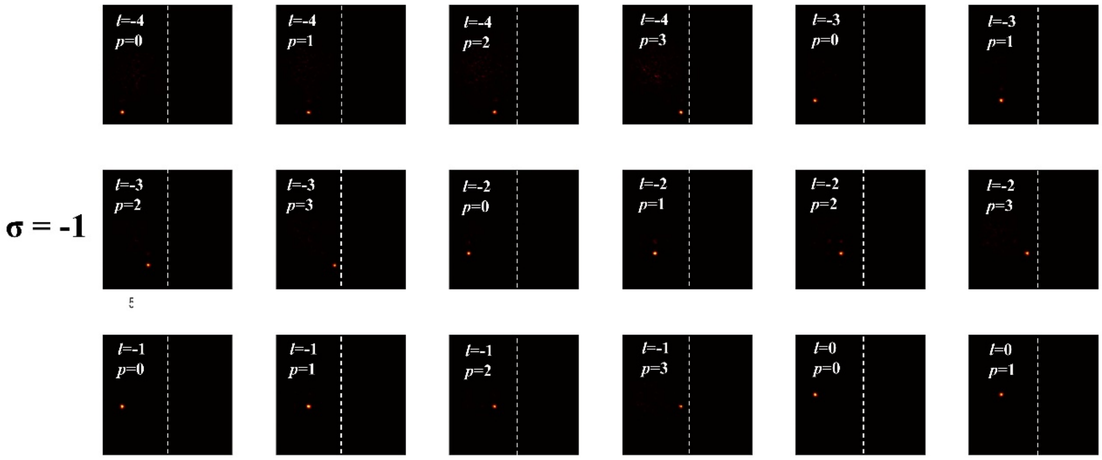

Fig. S1 Intensity distributions of LCP output channels from *l*=-4 to *l*=0 and *p*=0 to *p*=3.

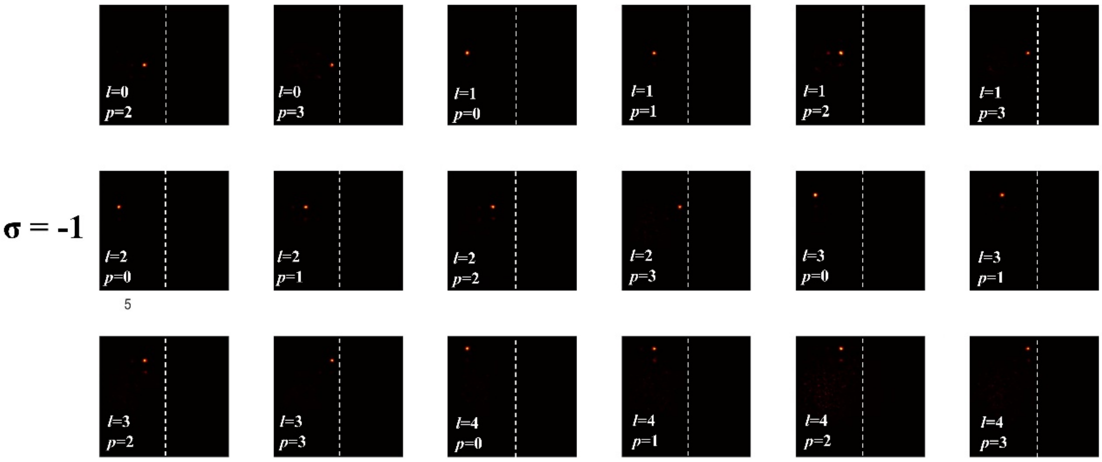

Fig. S2 Intensity distributions of LCP output channel from *l*=0 to *l*=4 and *p*=0 to *p*=3.

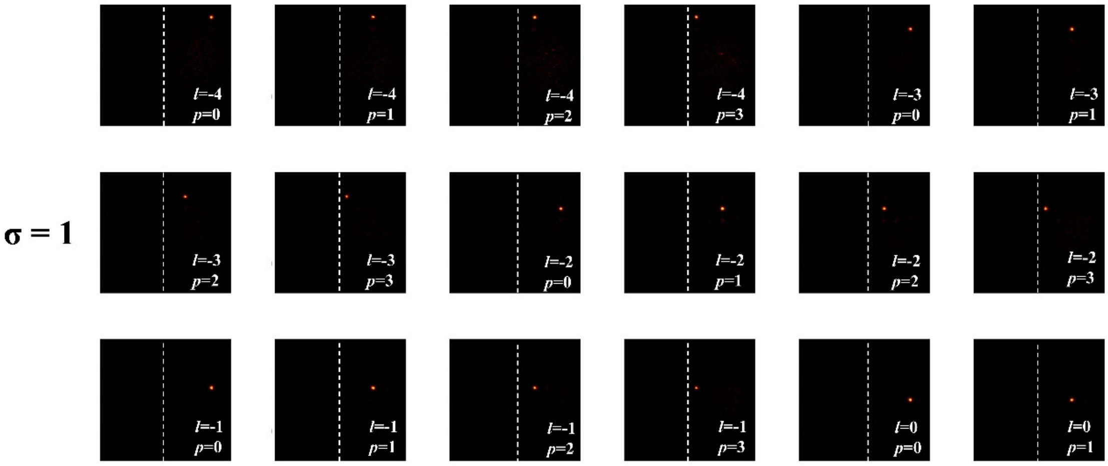

Fig. S3 Intensity distribution of RCP output channel from *l*=-4 to *l*=0 and *p*=0 to *p*=3.

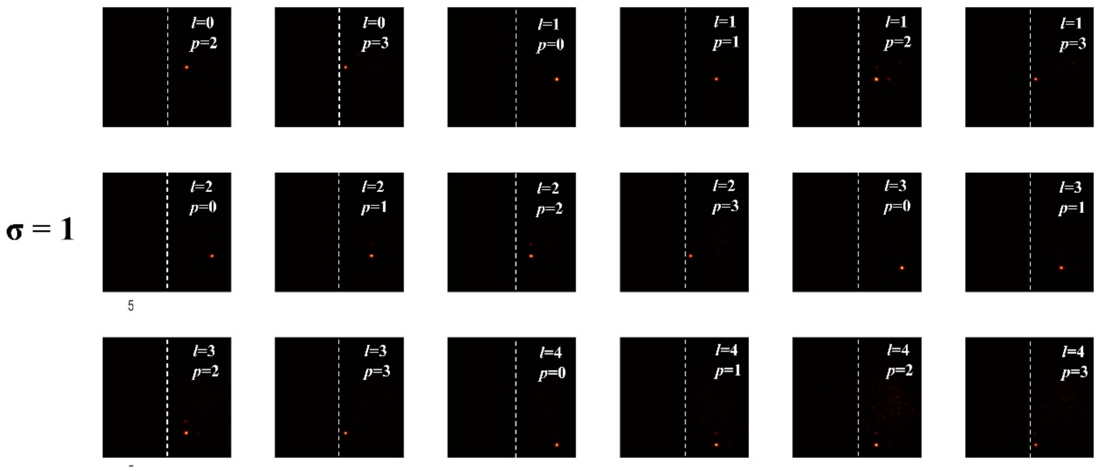

Fig. S4 Intensity distribution of RCP output channel from *l*=0 to *l*=4 and *p*=0 to *p*=3.

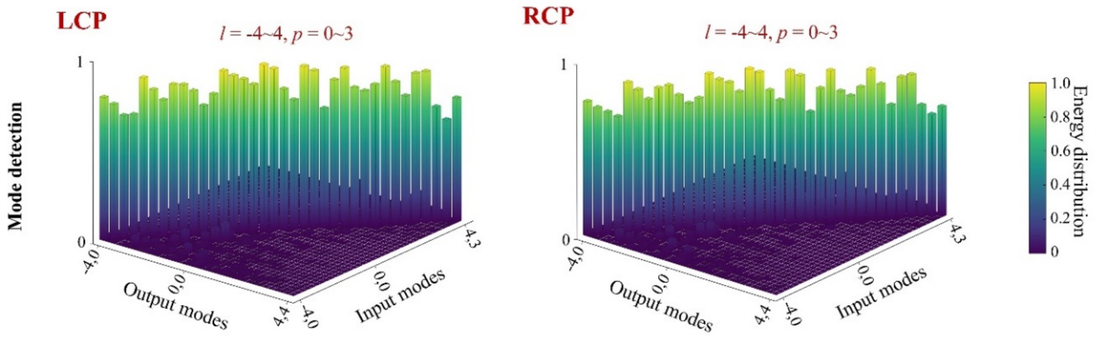

Fig. S5 The total energy ratio matrix of 36 modes for LCP and RCP.

Section IV: The Detailed results about patterns detection of superimposed VSBs.

We conducted an analysis to verify the mode detection results for both LCP and RCP. For the LG mode, we examined the azimuth index *l* = -2, -1, 1, 2, as well as the radial index *p* = 1, 2, 3. For the HG mode, we considered the index values of m = 1, 2, 3, 4 and n = 1, 2, 3. For the BG mode, we explored the azimuth index values *l* ranging from 1 to 12.

Using the identical approach, we calculated the energy ratio among the 36 channels on the output plane. Fig. S6 illustrates the results of this calculation. The energy ratio was determined by normalizing the light intensity of each channel through division by the total light intensity across all 36 channels.

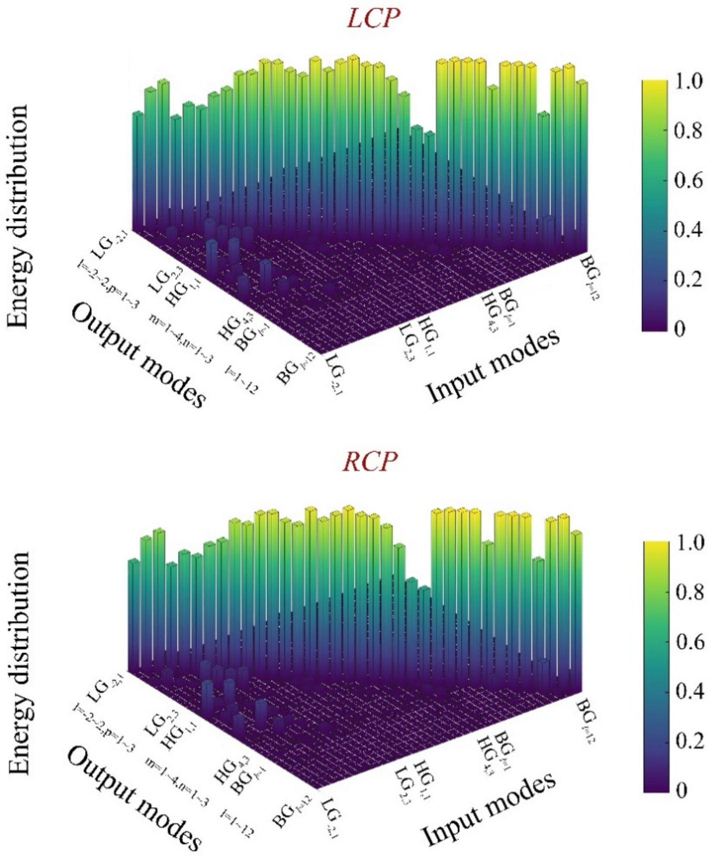

Fig. S6 The total energy ratio matrix of 36 modes for LCP and RCP.