Science Advances
AAAS

Supplementary Materials for

Lensless imaging with a programmable Fresnel zone aperture

Xu Zhang *et al.*

Corresponding author: Qian Chen, chenqian@njust.edu.cn; Chao Zuo, zuochao@njust.edu.cn.

Sci. Adv. **11**, eadt3909 (2025)
DOI: 10.1126/sciadv.adt3909

The PDF file includes:

Notes S1 to S10
Figs. S1 to S13
Legends for movies S1 to S4

Other Supplementary Material for this manuscript includes the following:

Movies S1 to S4

Note S1 Forward propagation model and reconstruction algorithm in traditional FZA-based lensless imaging

In the traditional Fresnel zone aperture (FZA) based lensless imaging system, the object $o(x, y)$ is illuminated by incoherent light. Each object point $p_{x,y}$ is projected onto the sensor plane with a distinct pattern $t(x, y)$, which corresponds to the magnification of the mask. Specifically, when the mask employed is the FZA, it can be expressed as:

$$t(x, y) = \frac{1}{2}\{1 + \cos[\beta(x^2 + y^2) + \phi]\}, \tag{S1}$$

or

$$t(r) = \frac{1}{2}[1 + \cos(\beta r^2 + \phi)]. \tag{S2}$$

Here, β and ϕ denote the coefficient and phase of the FZA, respectively, which are used to control the specific shape of the FZA. (x, y) are Cartesian spatial coordinates, and (r, θ) are the corresponding polar spatial coordinates. Considering the circular symmetry of FZA, the angular coordinates θ are omitted. Taking linear magnification m into account, the value of β is determined as $\beta = \beta_m/m^2$, where β_m is the FZA parameter designed on the mask. z_1 and z_2 denote the distances of object-mask and mask-sensor, respectively, from which the magnification $m = 1 + z_2/z_1$ can be obtained. The image $g(x, y)$ acquired by the sensor is a linear summation of the pattern for each object point p_o, which can be written as:

$$g(x, y) = \sum_{p_o \subset o(x,y)} p_o\, t(x, y) = o(x, y) \otimes t(x, y). \tag{S3}$$

In fact, in an incoherently illuminated scene, the reflectance and distance of different objects are encoded as the amplitude and phase of the equivalent wavefront in the coherent case. To reconstruct the equivalent complex amplitude, we utilize a four-step phase-shift sequence $\{0, \pi/2, \pi, 3\pi/2\}$ to change the phase of the FZA successively. Through the encoding process, we acquire the complex amplitude $g_{\text{comp}}(x, y)$ of the object $o(x, y)$:

$$g_{\text{comp}}(x, y) = \sum_{i=0}^{3} g_i(x, y) e^{-j\phi_i} = o(x, y) \otimes \exp[j\beta(x^2 + y^2)] = o(x, y) \otimes h(x, y), \tag{S4}$$

where $g_i(x, y)$ represents the captured image after modifying the phase ϕ_i of the FZA, and $h(x, y)$ is the equivalent complex point spread function. The elaboration of Eq. (S4) shows that we can achieve the incoherent-coherent equivalence process by encoding the intensity information of each incoherent point source as the coherent spherical wave with the same phase and different propagation distances. Once the complex amplitude $g_{\text{comp}}(x, y)$ is obtained, the amplitude information at different depths can be reconstructed by the virtual FZA method (*47*). The virtual FZA can be represented as:

$$h_v(x, y) = \exp[j\beta_v(x^2 + y^2)], \tag{S5}$$

where $\beta_v = \beta_m/(1 + z_2/z_1')^2$ is the parameter of the virtual FZA, and z_1' denotes the reconstruction depth. By observing Eq. (S4), the amplitude information can be acquired by performing the deconvolution operation in the frequency domain, which can be calculated as follows:

$$G(u, v) = \mathcal{F}\{g_{\text{comp}}(x, y)\} = \mathcal{F}\{o(x, y)\}\mathcal{F}\{\exp[j\beta(x^2 + y^2)]\}, \tag{S6}$$

$$G'(u,v) = G(u,v) \cdot \mathcal{F}\{h_v(x,y)\}^* = \mathcal{F}\{o(x,y)\}, \tag{S7}$$

$$I_{\text{comp}}(x,y) = \mathcal{F}^{-1}\{G'(u,v)\} = \mathcal{F}^{-1}\{\mathcal{F}\{o(x,y)\}\} = o(x,y). \tag{S8}$$

Here, $*$ indicates conjugate operation, $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ denote Fourier transform and inverse Fourier transform, respectively. (u,v) are the frequency domain coordinates corresponding to (x,y). I_{comp} is the complex amplitude at the distance z_1', at which distance objects will be reconstructed in sharp detail, while those far away will be blurred. By varying the value of z_1', we can reconstruct scene information at different depths.

Note S2 Comprehensive forward model of LIP considering the characteristics of the programmable FZA

In order to achieve high-quality lensless imaging with the integrated LIP module, a more detailed forward model considering the characteristics of the programmable FZA needs to be conducted. The first step is to reconsider the characteristics of the FZA patterns. To reconstruct the complex amplitude without twin image, a four-step phase-shift sequence $\{0, \pi/2, \pi, 3\pi/2\}$ is utilized to change the phase of the FZA successively, resulting in the equivalent complex amplitude H and equivalent transfer function T_{FZA}, as shown in Fig. S1A. It shows that we can achieve the incoherent-coherent equivalence process by encoding the intensity information of each incoherent point source as the coherent spherical wave with the same phase and different propagation distances. Furthermore, as we perform spatial shifting on the FZA within the fixed aperture, the density of FZA fringes also changes. After four-step phase shifting, the equivalent phase of the FZA pattern similarly experiences a shift. It can be seen that the offset α of the original FZA in the spatial domain after the four-step phase shifting is equivalent to the phase shift of H, while in the frequency domain corresponds to the translation of T_{FZA}. The ratio of the spatial and frequency domain offsets is $-\beta/\pi$.

Then, the frequency response characteristics of the programmable mask also impact the imaging process. The arrangement of pixels in a typical commercial liquid crystal display (LCD), as illustrated in Fig. S1B, reveals that the actual display area does not reach 100% due to the presence of driving circuits. Notably, as the pixel precision of the LCD increases, the augmentation of the fill factor faces heightened challenges. This, in turn, introduces additional influencing factors into our lensless imaging process, including the discrete sampling of patterns and pixel diffraction. Under the discrete pixel sampling of the LCD, the transfer function $T(u,v)$, induced by the pattern $t(x,y)$ displayed on the programmable mask, periodically replicates in the frequency domain with intervals of $[1/\Delta_x, 1/\Delta_y]$. When $T(u,v)$ is a band-limited function, and each replicated spectrum remains non-overlapping, accurate extraction of the complete spectrum can be achieved through band-pass filtering. This characteristic enables the precise restoration of the ideal transfer function $T(u,v)$. However, when considering the pixel size of LCD, the transfer function of the programmable mask $T_s(u,v)$, is constrained by a two-dimensional sinc function. As shown in the bottom of Fig. S1B, its primary spectral energy is confined within the first-order zero points, namely within $(2/a \times 2/b)$, while the spectrum beyond this range is predominantly suppressed and challenging to recover during the actual imaging process. In other words, the fundamental limitation to the resolution of lensless imaging based on the programmable mask stems from the pixel size rather than the pixel pitch.

Finally, by combining the forward models of the FZA pattern and the programmable masks, we can derive the frequency response characteristics of the lensless imaging system based on the programmable FZA, as shown in Fig. S1C. It can be observed that the equivalent transfer function T_{FZA} of the FZA, periodically replicates in the frequency spectrum with intervals of $[1/\Delta_x, 1/\Delta_y]$. As the FZA parameter β increases, each replicated spectrum also enlarges until they overlap with each other. In the spatial domain, this manifests as the periodic replication of the point spread function (PSF) and the mutual overlap in the reconstructed image. This aliasing problem was not mentioned in previous works using spatial light modulators (SLMs), mainly because programmable masks were not well modeled and exploited for their frequency response properties.

However, it is important to note that despite overlap, this does not imply the absence of high-frequency information within the central transfer function. When we further reduce the pixel size of the mask, the cutoff range of the two-dimensional sinc's zero points $2/a \times 2/b$ will expand. This is counter-intuitive, as we typically tend to equate pixel pitch with pixel size or solely consider pixel pitch, presuming that reducing pixel pitch is the only method to enhance resolution by recovering high-frequency information before aliasing.

Note S3 Offset-FZA parallel merging method based on Fourier ptychography

Traditional Fourier ptychography (FP) is a widely employed technique in microscopic imaging (*62, 65 and 66*), wherein multiple low-resolution images are synthesized from various illumination angles to translate the object's spectrum within the aperture. FP allows for phase recovery, high-resolution reconstruction and correction of phase aberrations. Drawing from this concept, we devised a process to achieve offset-FZA parallel merging in the programmable FZA-based lensless imaging system, as illustrated in Fig. S2. In addition, we utilized the concept of a difference map (*70-73*), in ptychography and process multiple algorithmic steps (Step 2, 4 and 5) in parallel to increase processing speed. The detailed algorithm flow is outlined below.

Step 1: Images capture under optimal parameter matching. The images $g_{i,j}(x, y)$ are acquired under $\mathrm{Mask}_{i,j}$ with different phase $\phi = \{0, \pi/2, \pi, 3\pi/2\}$ and the optimal parameter β_m and R_m, where i, j is the offset of the mask relative to the central aperture.

Step 2: Parallel reconstruction with four-step phase shifting. Four measurements of each sub-aperture are utilized to reconstruct the complex amplitude $g_{\mathrm{comp}\ i,j}$ with the four-step phase-shift algorithm demonstrated in Note S1. Meanwhile, according to the spectral range and spectral offset given in the main text, a frequency domain mask $P_{i,j}$ centred at f_{step} and with f_{diam} as the diameter can be obtained to perform spectrum constraints on the sub-aperture reconstruction results. The decoupling of information between each sub-aperture enables parallel reconstruction of intensity information.

$$I_{i,j} = \left|\mathcal{F}^{-1}\{P_{i,j} \cdot \mathcal{F}\{g_{\mathrm{comp}\ i,j}\}\}\right|^2. \tag{S9}$$

Step 3: Initial spectrum guess. The Fourier transform of the central aperture intensity $I_{0,0}$ and the extracted reconstructed spectrum range P_{init} are used to obtain the initialized target spectrum U_0.

$$U_0 = P_{\text{init}} \cdot \mathcal{F}\{I_{0,0}\}. \tag{S10}$$

Step 4: Sub-spectrum extraction in parallel. Based on the spectral range and spectral offset given in the main text, the spectrum $U_{i,j}^e$ corresponding to the reconstruction result under the sub-aperture is extracted from U_0 by $\text{Mask}_{i,j}$ in parallel, and the complex amplitude $u_{i,j}^e$ at this position can be obtained by inverse Fourier transform.

$$U_{i,j}^e = U_0 \cdot \text{Mask}_{i,j}, \tag{S11}$$

$$u_{i,j}^e = \mathcal{F}^{-1}\{U_{i,j}^e\}. \tag{S12}$$

Step 5: Spatial intensity constraints in parallel. Using the intensity information $I_{i,j}^c$ reconstructed from the practical data captured at this offset, the intensity component $I_{i,j}^e$ of the extracted sub-aperture corresponding to the complex amplitude $u_{i,j}^{e\,\prime}$ is updated with the following equation in parallel:

$$u_{i,j}^{e\,\prime} = u_{i,j}^e \sqrt{I_{i,j}^c / I_{i,j}^e}. \tag{S13}$$

Step 6: Spectrum global update. The spectrum of all sub-apertures $u_{i,j}^e$ is weighted and merged, and the spectrum results U_k of kth round are updated in step size α:

$$U_{\text{sum}}^e = \sum_{i,j} \mathcal{F}\{u_{i,j}^{e\,\prime}\} P_{i,j}(u,v), \tag{S14}$$

$$P_{\text{sum}} = \sum_{i,j} P_{i,j}(u,v), \tag{S15}$$

$$U_k = (1-\alpha)U_{k-1} + \alpha \frac{U_{sum}^e}{P_{\text{sum}}+\delta},\ k = 1, \dots, n, \tag{S16}$$

where δ represents the regularization parameter. U_{sum}^e is the summation of the sub-aperture spectrum, and P_{sum} is the summation of frequency-domain masks $P_{i,j}$, respectively.

Step 7 & 8: Convergence criterion. Repeat Steps 4 to 6 until the complex amplitude converges or the update rate is less than the set threshold to obtain the optimal solution U_{OPM} for the offset-FZA parallel merging complex amplitude reconstruction I_{OPM} and ϕ_{OPM}.

Note S4 Simulations under different pixel sizes and β

The presence of aliasing does not necessarily result in the complete loss of high-frequency information, which should be duly noted. For an imaging system, as long as it can respond to incoming high-frequency information, the information aliased within the cutoff frequency has the potential for complete recovery. Based on the analysis of the programmable FZA imaging system's response, the true limiting factor for its upper-frequency response is the pixel size rather than the pixel pitch.

Simultaneously, for most programmable masks, their fill factors do not reach 100%, implying that the cutoff frequency caused by the pixel size mentioned earlier will generally be higher than the replication frequency brought about by the pixel pitch. Therefore, even with aliasing present, high-frequency information can still be reconstructed. As illustrated in Fig. S3A, we simulated the reconstruction results under a 5× aliasing scenario. It is noticeable that with an increase in β, aliasing artifact occurs both in the frequency domain and spatial domain of the reconstruction results. However, if we maintain a constant pixel pitch and only decrease the pixel size, the central spectrum range limited by the zero points of the two-dimensional sinc function is further extended. From the perspective of the PSF, as shown in Fig. S3B, under the same β parameter, the reconstruction results with smaller pixel sizes correspond to narrower PSFs, indicating that despite aliasing, the enhancement of high-frequency details is still achieved.

Note S5 Reconstruction through optimal parameter matching and offset-FZA parallel merging

To achieve aliasing-free, high-quality lensless reconstruction based on programmable FZA, we have applied optimal parameter matching and the offset principle of FZA to the process of offset-FZA parallel merging reconstruction. As depicted in Fig. S4A, we conducted simulations to validate the previously introduced optimal parameter matching, comparing the impacts of optimal β, smaller β, and larger β on the outcome of offset-FZA parallel merging reconstruction.

It is evident that a smaller β results in sub-optimal utilization of spectral space, leading to decreased reconstruction efficiency. Meanwhile, the lower overlap rate makes it challenging to meet the requirements of spectrum synthesis. Conversely, a larger β introduces aliasing in the information from each aperture, restricting the quality of offset-FZA parallel merging reconstruction. Under our proposed optimal sampling criterion, both the central aperture and the surrounding apertures are unaffected by the periodically replicated spectrum, ensuring high-quality reconstruction in each sub-aperture. This is pivotal for achieving aliasing-free, high-quality frequency domain synthesis. Additionally, maximizing spectrum utilization aids in both collection efficiency and aperture overlap rate, enabling us to achieve high-quality reconstruction with a relatively small number of apertures during frequency domain synthesis.

In Fig. S4B, we compared the PSF before and after spectrum synthesis under different β values. It is observable that the PSF synthesized with a smaller β exhibits significant sidelobe peaks. Inversely, both optimal β and larger β avoid this issue due to a higher overlap rate. However, the PSF synthesized with a larger β, while having a similar Full Width at Half Maximum (FWHM) to our chosen optimal β, demonstrates a slow decay in the sidelobes at the bottom, which has some impact on the reconstruction quality.

Note S6 Image SNR evaluation method without the reference image

The signal-to-noise ratio (SNR) is defined as the ratio of the signal strength to the noise level. In the era of digital imaging, every photo we captured is obtained through a light-sensitive element, resulting in inherent noise within our images. Consequently, when calculating the SNR for an image, it is essential to separate and analyze these signal and noise elements individually. Traditionally, this can be achieved by obtaining a reference image as ground truth and another image specifically for SNR calculation purposes. The computation method for determining SNR can be described as follows:

1) Subtract the reconstructed image A to be computed from the original image B to an image C ($A = B + C$);
2) Consider B as the signal part of A and C as the noise part of A;
3) Calculate the variance of B and C respectively;
4) Calculate the ratio of the two items above to get the value of SNR.
5) Logarithmize the value of SNR to get the value of SNR in dB.

For the calculation of SNR of the reconstructed image without reference image, it is necessary to separate the signal and noise in the image, namely the foreground and background. In general, the foreground is the signal and the background is the noise. Therefore, we can select a reconstructed image with clear black and white as the evaluation basis of SNR (for example, USAF resolution target), select the white area as the superposition area of signal and noise (A), and the black area as the noise area (C), and then calculate the SNR using the above method. It is worth noting that, because the white area and the black area are not completely overlapped in space, the SNR calculated by this method is not an absolute value, and is only applicable to the SNR comparison of different methods in the same scene.

Note S7 Considerations for raw data during the measurement process

The sensor's measurements are the only indirect raw data we can obtain, so it is necessary to discuss the specific conditions encountered during data collection. Thus, we analyze the data from the perspectives of sensor measurement overflow and the characteristics of color and spatial information.

1) Sensor Measurement Overflow: In our explanation of the OPM method, we noted that each pattern's sampling must be considered when configuring lensless imaging system parameters (Fig. S7A). When system parameters are configured according to the optimal parameter matching method we proposed (Fig. S7B), all patterns can be fully sampled, meaning each point-source response is completely recorded, without loss of frequencies, which is a factor essential for high-quality reconstruction. However, if the relationship between aperture size and sensor size does not satisfy our parameter selection criteria, some patterns will not be fully sampled, resulting in frequency losses in edge information and causing ringing artifacts (Fig. S7C). Additionally, if the aperture size is reduced to allow complete sampling of the patterns, FZA parameter selection becomes a further consideration. If the β parameter of the FZA is too small, it limits the use of the LCD screen's frequency response, resulting in a loss of imaging resolution (Fig. S7D). Conversely, when FZA parameters are too large, the inherent physical properties of the LCD screen limit imaging quality, leading to reduced SNR (Fig. S7E). Consequently, our proposed OPM method allows optimal determination of aperture size and FZA parameters based on the physical parameters of the LCD screen and sensor, enabling higher-quality lensless reconstructions..

2) Characteristics of Color and Spatial Information: In our experiments, we used a color sensor as the key component for measuring the encoded images, which necessarily entails balancing color and spatial information (Fig. S7F). When using a grayscale sensor to measure and reconstruct an object, each pixel only captures the variation in spatial light intensity, so reconstruction can be carried out directly according to the proposed OPM method without considering color information, as illustrated in Fig. S7 (G and H). In contrast, when imaging color objects, it is necessary to demosaic and separate the collected raw images into RGB channels, apply the OPM method to each channel, and finally combine the three reconstructed channels into a color image, as shown in Fig. S7 (I and J). During the process, the use of an R-G-G-B Bayer filter reduces the sensor's native spatial resolution, but the color and spatial information has minimal impact on reconstruction due to two primary factors: first, current demosaicing algorithms for color sensors are very mature, and the test shows that the built-in algorithm of an industrial camera is sufficient for our experimental needs (*59*); second, the analysis of the imaging system's resolution shows that, due to the significantly larger pixel size of the LCD screen compared to that of the sensor, no pixel aliasing occurs with the current experimental parameters.

In conclusion, the proposed OPM method addresses the two specific conditions encountered during sensor measurement, providing a solid foundation for subsequent high-quality lensless reconstructions.

Note S8 Transmission rate of the programmable LCD

Due to the inherent polarization characteristics of commercial LCDs, the optical efficiency utilized by our LIP module is limited. Testing shows that the LCD used has a transmission rate of 40% at 550 nm. This issue is common to all spatial light modulator devices based on polarization modulation. However, our method mitigates this limitation by exploiting complementary information across multi-frame measurements, achieving high-quality reconstructions even under constrained sensor SNR conditions. Fig. S8 shows the PSNR reconstruction curves for different noise levels, indicating that even with increased noise variance, our method achieves stable and robust PSNR performance in both static and dynamic modes. This suggests that despite the reduced optical flux, our OPM method still delivers satisfactory reconstruction quality. In future work, we plan to explore modulators with higher transmittance to further unlock the potential of the LIP module.

Note S9 Comparison of LIP with state-of-the-art lensless imaging methods

To further demonstrate the robustness of our LIP method compared to the traditional single-shot lensless imaging methods, we chose 10 high-quality, intricate-colored objects from the DIV2K dataset (*74*) and conducted simulations with different zero mean Gaussian noise levels (variances σ = 1e-7, 1e-6, and 1e-5), as shown in Fig. S10. Each mask's forward imaging process can be modeled as an optimization problem. Therefore, when solving the inverse problem for each mask, we used Two-Step Iterative Shrinkage/Thresholding (TwIST) (*19, 23*) as the optimization method and total variation (TV) regularization as the prior constraint. For the traditional static-modulation reconstruction method, the optimal regularization parameter τ is actually dependent on the optimization algorithm, the scene, and the noise level, which is impractical to always guarantee the selection of the optimal parameter τ. In the simulation, we set the regularization parameter τ to 0.006 to ensure that each algorithm gives a practical performance, and the parameter also provides reliable measurement data.

We first selected Object 1 as the simulation target for our method in dynamic mode (N = 4) and traditional static-modulation methods. The reconstruction results are shown in Fig. S10A. For a fair comparison, we simulated the reconstruction results of the traditional methods for single-frame (N = 1) and averaged multi-frame average (N = 4) measurements. At the same noise level, our method significantly outperforms the traditional single-frame (N = 1) reconstruction method regarding imaging quality, color crosstalk, and noise suppression. Regarding multi-frame (N = 4) results, as the noise variance σ increases, the inflexibility of traditional static-modulation methods becomes apparent, making it difficult to counteract the effects of noise simply by averaging multiple frames. In contrast, leveraging the complementary nature of different masks in the programmable FZA, our method shows excellent noise resistance, with no significant decline in reconstruction quality across three different noise levels. Next, we selected Object 2 as the simulation target for our method in static mode (N = 16) and traditional static-modulation methods. The reconstruction results are shown in Fig. S10B. Compared to the dynamic mode, reconstruction artifacts in the static mode are greatly suppressed, and the overall imaging quality significantly improves. However, traditional static-mask methods with N = 16 measurements remain highly sensitive to noise, consistent with the conclusions in Fig. S10A. Fig. S10C and Fig. S10D show the PSNR comparison curves of dynamic/static mode and traditional static-mask lensless imaging (N = 4 / N = 16) under different noise variances. When the noise variance σ is 1e-5, our method in the dynamic mode achieves a PSNR of 13.74 dB higher than the best reconstruction result of the Contour mask with the traditional method and 9.80 dB higher in the static mode. The above simulation results indicate that our method can utilize the complementary information between different patterns modulated by the programmable mask, achieving higher quality and better robustness to noise than traditional static-modulation methods, even with the same number of measurements.

Then, compared to the traditional multi-shot lensless imaging methods, our FZA-based lensless imaging approach offers distinct advantages. While recent studies have explored dynamic masks, such as random mask array (*44*) and translated separable mask (*36*), to broaden imaging functions and enhance quality, these methods inevitably face several challenges during measurement and reconstruction:

1) Requirement for High-Precision Calibration: For non-analytical multi-shot masks, ensuring alignment between the designed and manufactured patterns often proves challenging, necessitating precise calibration before the experiment. Furthermore, recalibration becomes unavoidable if external factors like prolonged vibration or structural aging alter the mask-to-sensor distance. As the number of masks increases, so too do the complexity and data volume for calibration, thereby reducing reconstruction efficiency.

2) Uncertainty in Mask Frequency Response: Ideally, multi-shot methods should employ well-designed masks that remain orthogonal in the transform domain to maximize imaging efficiency and quality. However, many existing methods, such as (*44*), adopt random multi-mask lensless imaging systems, resulting in limited interpretability and orthogonality among masks in the frequency domain. The lack of orthogonality leads to overlapping information and reduced acquisition efficiency.

3) Requirement for Prior Knowledge about the Target Object: Traditional multi-shot lensless imaging methods model the optimization problem similarly to single-shot approaches, with the additional data fidelity term to strengthen constraints. A classical approach to addressing this ill-posed inverse problem is to impose certain prior assumptions on the solution, reformulating the original ill-posed problem as a well-posed optimization problem to facilitate regularization. To ensure convergence, it is often necessary to apply regularization priors to the target object. For instance, TV regularization assumes that the object is block-smooth, introducing sparse priors in the gradient domain (*19*). However, this assumption lacks universal applicability and robustness for arbitrary objects, and performance may degrade for complex objects that do not exhibit block-smooth characteristics.

In contrast, our programmable FZA-based lensless imaging system leverages the unique properties of the FZA to address these challenges:

1) Analytical Form of the FZA: Since the FZA serves as a diffraction-interference element, the captured pattern images align with holographic interference patterns in an analytical form, thus eliminating the need for precise PSF calibration.

2) Frequency Domain Orthogonality of Offset FZA: Fourier domain analysis of the FZA after a four-step phase shifting reveals that its spectrum is confined to a limited range, unlike other masks that distribute across the entire frequency domain, as illustrated in Fig. S9. By shifting the FZA, its frequency range can be adjusted accordingly, allowing precise control over the mask's frequency response range. The feature allows for the possibility of spectral synthesis. When appropriate parameters are selected, the frequency ranges of the shifted and unshifted FZAs remain non-overlapping in the dynamic mode, demonstrating that FZA can achieve frequency domain orthogonality, which is critical for maximizing acquisition efficiency.

3) Phase Retrieval Property of FZA: As previously mentioned, the FZA pattern resembles the point-source interference pattern seen in incoherent self-interference holography, such as FINCH (*48*). The similarity suggests that once the phase of the captured image is retrieved, refocusing at various depths becomes possible. More importantly, once the phase is retrieved, we can apply coherent optical system theory to analyze what is otherwise an incoherent lensless imaging system. By leveraging the frequency offset characteristic introduced by the offset FZA, we can integrate the imaging process with ptychograghy imaging, achieving a high-resolution, high-SNR quasi-coherent lensless reconstruction through frequency domain synthesis of offset FZA sub-apertures.

Finally, we qualitatively compare lensless imaging approaches that include static masks, dynamic mask-based lensless imaging, LIP in dynamic mode, and LIP in static mode across six aspects (resolution, imaging speed, field of view (FoV), cost-effectiveness, calibration robustness, and flexibility) in Fig. S12, to illustrate the advantage of our method:

1) Resolution and Imaging Speed: Static mask-based lensless imaging typically requires incorporating regularization priors to solve this ill-posed optimization problem. As a result, when noise increases or the model mismatches, high-frequency information of the object buried in noise cannot be effectively retrieved, making it difficult to achieve high-resolution, high-SNR reconstructions from single-shot measurement. In contrast, dynamic mask-based methods and our LIP can extract complementary frequency information from multiple modulated measurements, improving reconstruction quality. In static mode, LIP can precisely extract and synthesize information at different frequencies, supported by an accurate forward model of the programmable mask and FZA, achieving high-resolution, high-SNR reconstructions with adjustable imaging speed and quality.

2) FoV, Cost-effectiveness, and Calibration Robustness: The primary limiting factor stems from the characteristics of the mask. Static masks, whether lithographic masks (separable mask, contour mask) or random masks (diffusers, scotch tapes, blood cell slides), can offer large FoV and acceptable cost but require precise PSF calibration (for different depths and mask distributions) before use. Dynamic masks, on the other hand, typically use scientific-grade spatial light modulators (SLMs), which have limited FoVs and higher costs and also require precise calibration of non-analytical masks. However, in our LIP module, we leverage a commercial large-FoV LCD (±40°), produced at scale, to keep costs controlled. By combining the discrete FZA forward model with the OPM algorithm, we achieve precise scene reconstructions without the need for calibration.

3) Flexibility: Traditional static mask-based lensless imaging is application-specific, requiring new mask designs for different scenarios. Conversely, dynamic mask-based systems allow for easy adaptation of the mask pattern according to the application needs. By modulating across the time domain, dynamic masks can capture multiple encoded measurements at different time points, thus increasing the dimensionality of the information. This enables the system to simultaneously obtain images under various masks, effectively improving spatial, spectral, or depth resolution and facilitating multi-functional and adaptive imaging within one system.

Note S10 The connection between FZA and incoherent self-interference holography

The FZA pattern essentially corresponds to the point-source interference pattern observed in incoherent self-interference holography (*48*). In a typical incoherent self-interference digital holographic system based on a Michelson interferometer, the two wavefronts that interfere in the sensor plane can be written as u_1 and u_2. Here, u_1 represents the plane wave of the unaltered object, while u_2 is a spherical wave formed by reflection through a concave mirror. These wavefronts can be mathematically described as follows:

$$u_1 = \exp(jkz), \tag{S17}$$

$$u_2 = \exp(jkz)\exp[j\frac{k}{2z}(x^2+y^2)], \tag{S18}$$

where $k = 2\pi/\lambda$ is the wave number, and (x, y, z) represents the spatial coordinate. Ignoring the constant phase factor $\exp(jkz)$ in both wavefronts, the resulting intensity pattern produced by their interference on the sensor can be expressed as:

$$I_{\text{inf}} = |u_1 + u_2|^2 = \langle u_1\rangle^2 + \langle u_2\rangle^2 + 2\langle u_1, u_2\rangle = I_1 + I_2 + \cos\left(\frac{k}{2z}r^2\right). \tag{S19}$$

Comparing I_{inf} with the numerical form of the FZA t_{FZA},

$$t_{FZA} = \frac{1}{2}[1 + \cos(\beta r^2 + \phi)]. \tag{S20}$$

It can be observed that the expression for the FZA pattern, t_{FZA}, aligns with the interference pattern I_{inf} of incoherent self-interference digital holography, both taking the form of a quadratic cosine function, as shown in Fig. S13. By adjusting the parameter β, FZA patterns can be accurately generated at any desired depth. The distinctive properties of the FZA provide a robust theoretical foundation for lensless imaging under incoherent illumination, drawing parallels to incoherent self-interference holography. This enables depth refocusing by employing a four-step phase-shift method to capture depth-related phase information from the measurement image. Moreover, the depth-dependent characteristic of FZA-based lensless imaging eliminates the traditional need for precise PSF calibration at each depth, significantly simplifying the imaging process.

Compared to the strong connection between FZA-based lensless imaging and holography, traditional lensless imaging simply models the imaging process as a convolution of the object intensity with a designed mask pattern, requiring deconvolution of the measurements to solve the inverse problem. This imaging and solution model relies more on complex mathematical optimization, overlooking the support provided by the actual physical imaging process, thus necessitating extensive PSF calibration to bridge this gap.

On the other hand, the formation of the FZA pattern can also be understood from the perspective of the beam-split interference theory of FZA. By expressing t_{FZA} in its exponential form, we can derive:

$$\begin{aligned} t_{FZA} &= \frac{1}{2}[1 + \cos(\beta_0 r^2 + \phi_i)] \\ &= \frac{1}{2}\left[1 + \frac{1}{2}\exp(j\beta_0 r^2 + \phi_i) + \frac{1}{2}\exp(-j\beta_0 r^2 - \phi_i)\right] \\ &= \frac{1}{2} + \frac{1}{4}\exp(j\beta_0 r^2 + \phi_i) + \frac{1}{4}\exp(-j\beta_0 r^2 - \phi_i)\,. \end{aligned} \tag{S21}$$

When parallel light is incident on the FZA, it can be divided into three beams: The first term is the DC component, with an intensity equal to half of the incident light; the second term is the divergent light, with an intensity equal to one-quarter of the incident light and a phase in the form of a positive quadratic term; the third term is the convergent light, with an intensity equal to one-quarter of the incident light and a phase in the form of a negative quadratic term. Ignoring the effect of the quadratic phase, the interference intensity of the three beams is:

When parallel light is incident on the FZA, it splits into three distinct beams:

1) The **DC component**, with an intensity equal to half of the incident light.

2) The **divergent beam**, with an intensity equal to one-quarter of the incident light and a phase characterized by a positive quadratic term.

3) The **convergent beam**, with an intensity equal to one-quarter of the incident light and a phase characterized by a negative quadratic term.

By ignoring the constant and the effect of the quadratic phase term, the interference intensity of these three beams can be expressed as:

$$I_{\mathrm{tri}} = \frac{1}{2}\cos\left(\beta_0 r^2 + \phi_i\right) + \frac{1}{8}\cos\left(2\beta_0 r^2 + 2\phi_i\right). \tag{S22}$$

From the above equation, it can be concluded that using the FZA to encode parallel light generates ring-shaped interference fringes in the light field. This pattern is consistent with the interference pattern produced by a point source and a plane wave under coherent illumination in digital holography. This observation suggests that the FZA has the capability to encode incoherent light fields into mixed coherent light fields, thus providing a theoretical foundation for the subsequent reconstruction of incoherent light fields.

When we apply the four-step phase-shifting technique from traditional holography to the above equation, we obtain:

$$\begin{aligned}
I_{\mathrm{tri1}} &= \frac{1}{2}\cos(\beta_0 r^2) + \frac{1}{8}\cos(2\beta_0 r^2)\,, \\
I_{\mathrm{tri2}} &= -\frac{1}{2}\sin(\beta_0 r^2) - \frac{1}{8}\cos(2\beta_0 r^2)\,, \\
I_{\mathrm{tri3}} &= -\frac{1}{2}\cos(\beta_0 r^2) + \frac{1}{8}\cos(2\beta_0 r^2)\,, \\
I_{\mathrm{tri4}} &= \frac{1}{2}\sin(\beta_0 r^2) - \frac{1}{8}\cos(2\beta_0 r^2)\,.
\end{aligned} \tag{S23}$$

Then, we can obtain the complex amplitude I_{comp} under incoherent holography as follows:

$$\begin{aligned}
I_{\mathrm{comp}} &= [I_{\mathrm{tri1}} - I_{\mathrm{tri3}}] + j[I_{\mathrm{tri2}} - I_{\mathrm{tri4}}] \\
&= \cos(\beta_0 r^2) + j\sin(\beta_0 r^2) \\
&= \exp(j\beta_0 r^2).
\end{aligned} \tag{S24}$$

From the above equation, it can be concluded that applying the theory of three-beam interference from the FZA to explain the forward model of lensless imaging based on FZA, as discussed in previous chapters, yields consistent results. This consistency highlights the shared underlying principles between FZA-based lensless imaging and incoherent self-interference holography, while also emphasizing the differences from traditional lensless imaging methods that rely on fixed masks and numerical optimization reconstruction.

More importantly, once the phase information is retrieved, the incoherent lensless imaging system can be analyzed using the principles of coherent optical imaging systems. By integrating the proposed OPM method, we achieve calibration-free, high-resolution, and high-SNR quasi-coherent lensless reconstructions through the frequency domain synthesis of different offset FZA sub-apertures.

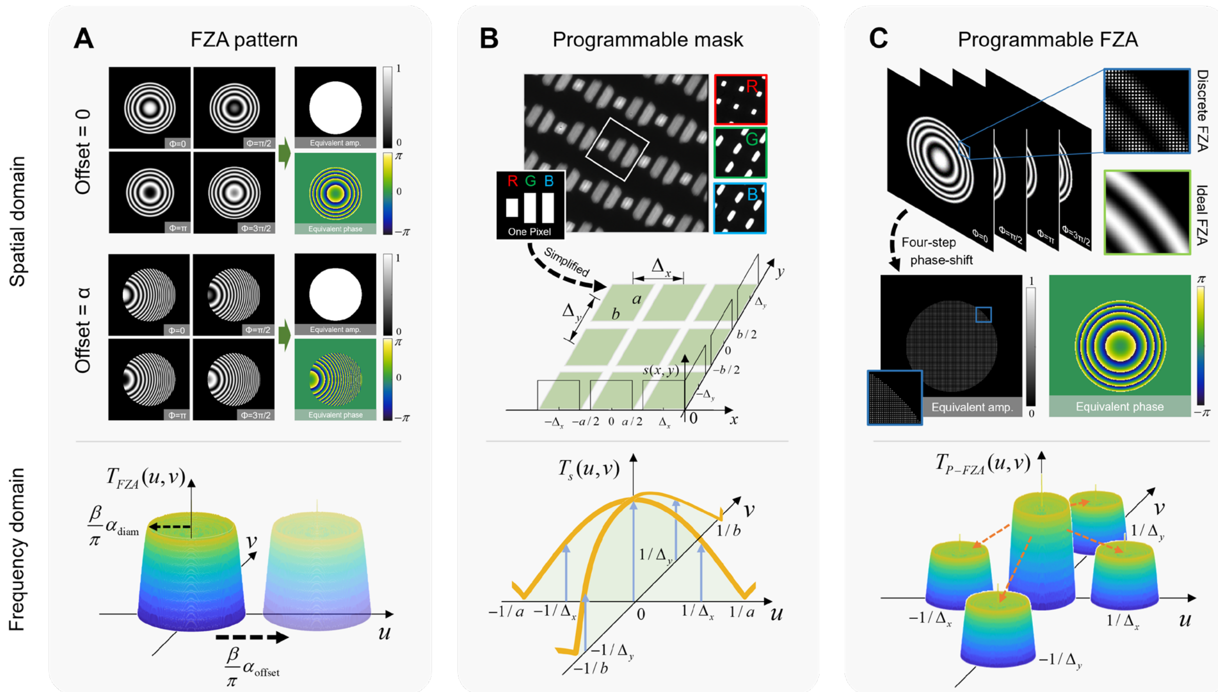

Fig. S1 The forward model of the lensless imaging system with programmable FZA.

The spatial and frequency domain characteristics of (**A**) FZA pattern, (**B**) programmable mask and (**C**) programmable FZA.

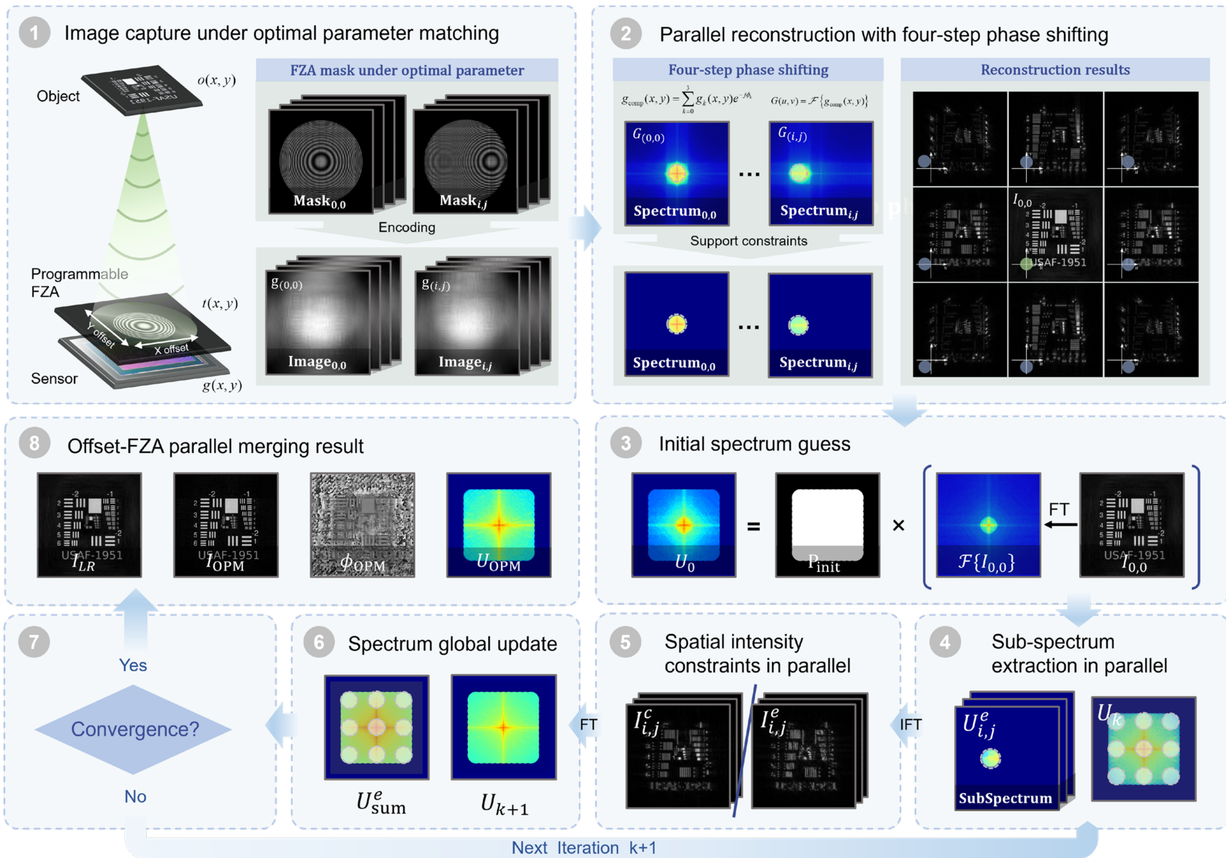

Fig. S2 The algorithm flow of the offset-FZA parallel merging method of LIP.

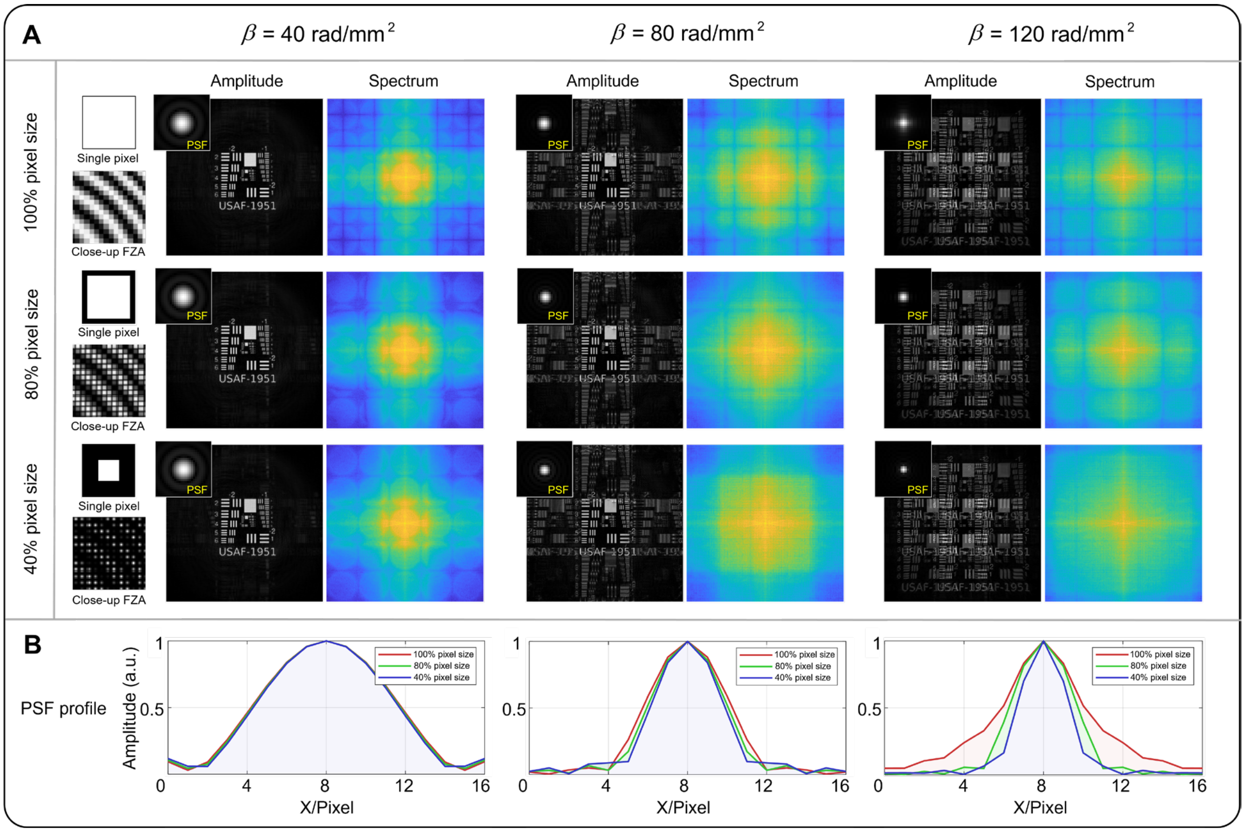

Fig. S3 Reconstruction results at different pixel sizes and $\boldsymbol{\beta}$.

(**A**) Reconstruction results at different LCD pixel sizes with β=40, 80, and 120 rad/mm^2; (**B**) Comparison of PSF profiles reconstructed at different LCD pixel sizes with β =40, 80, and 120 rad/mm^2.

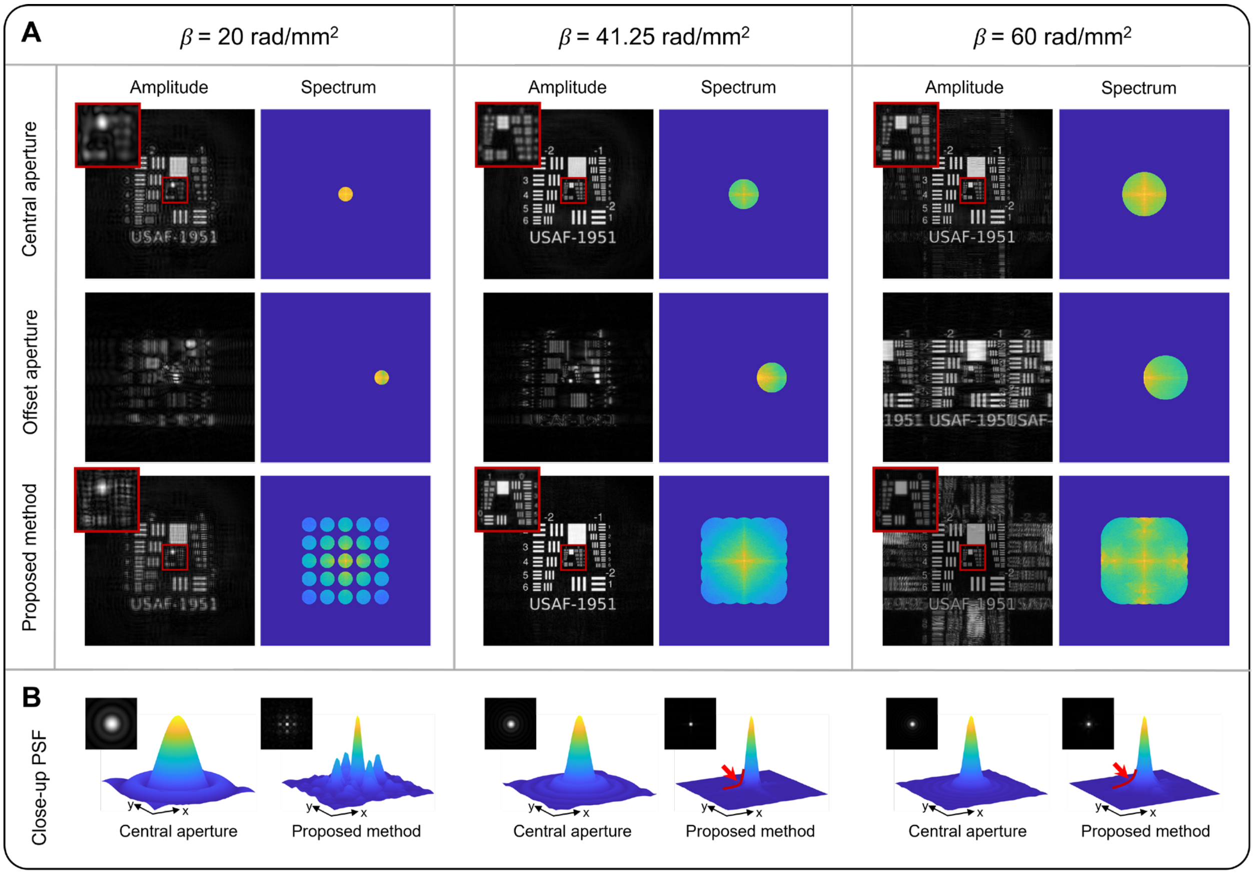

Fig. S4 Offset-FZA parallel merging reconstruction using the optimal parameter matching (with the USAF resolution target as the object).

(**A**) Reconstruction results of center aperture, offset aperture, and proposed method for β = 20, 41.25, and 60 rad/mm^2; (**B**) PSF comparison of center aperture with reconstruction from the proposed method for β = 20, 41.25, and 60 rad/mm^2.

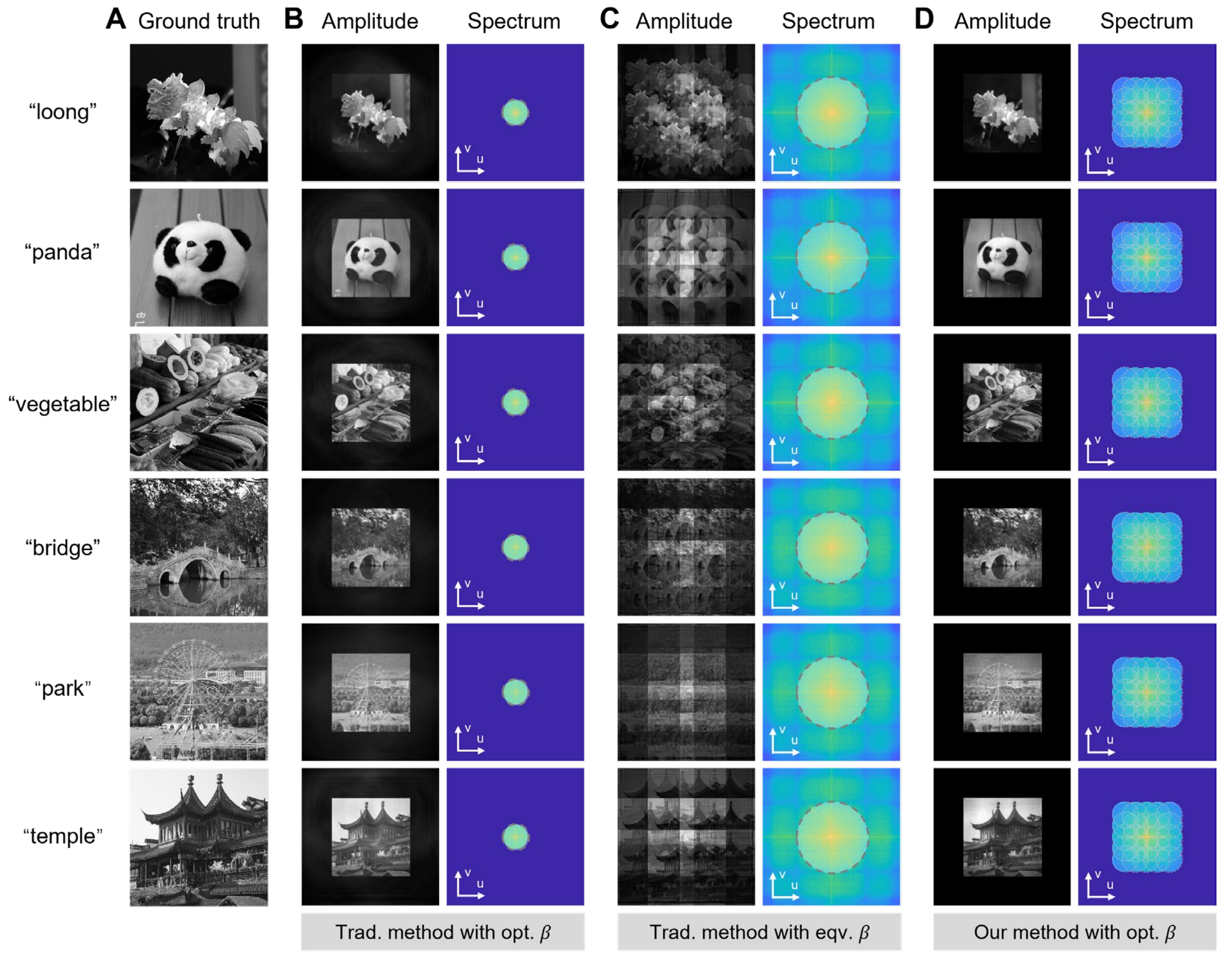

Fig. S5 Numerical simulations on various complex objects.

(**A**) Ground truth of Object "loong" ~ "temple." Reconstruction results of (**B**) traditional method with optimal β, (**C**) traditional method with equivalent β, and (**D**) our method with optimal β for Object "loong" ~ "temple." Photo credit: Xu Zhang.

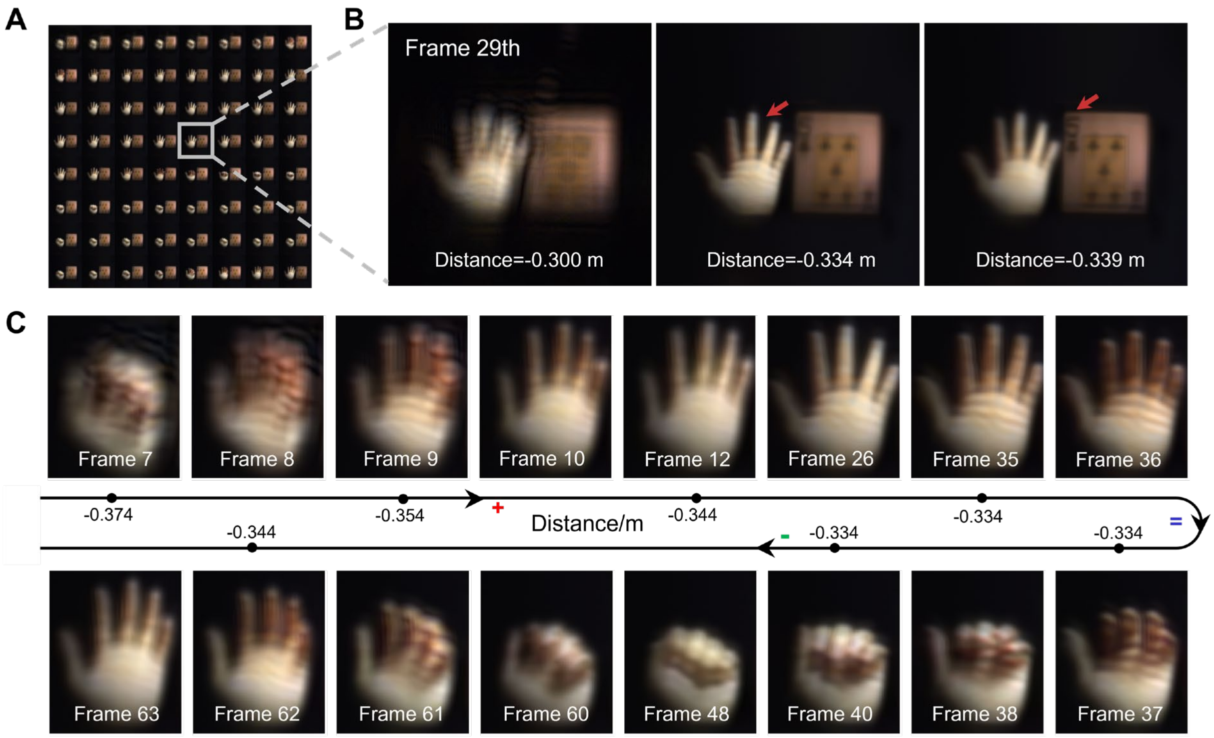

Fig. S6 Reconstruction results of a dynamic scene experiment.

(**A**) Captured and reconstructed dynamic results for a total of 64 frames, with the scene containing a playing card and a hand; (**B**) Refocusing results for Frame 29th at -0.300 m, -0.334 m, and -0.339 m, respectively; (**C**) Dynamic reconstruction of the hand, where the refocusing position varies near and far.

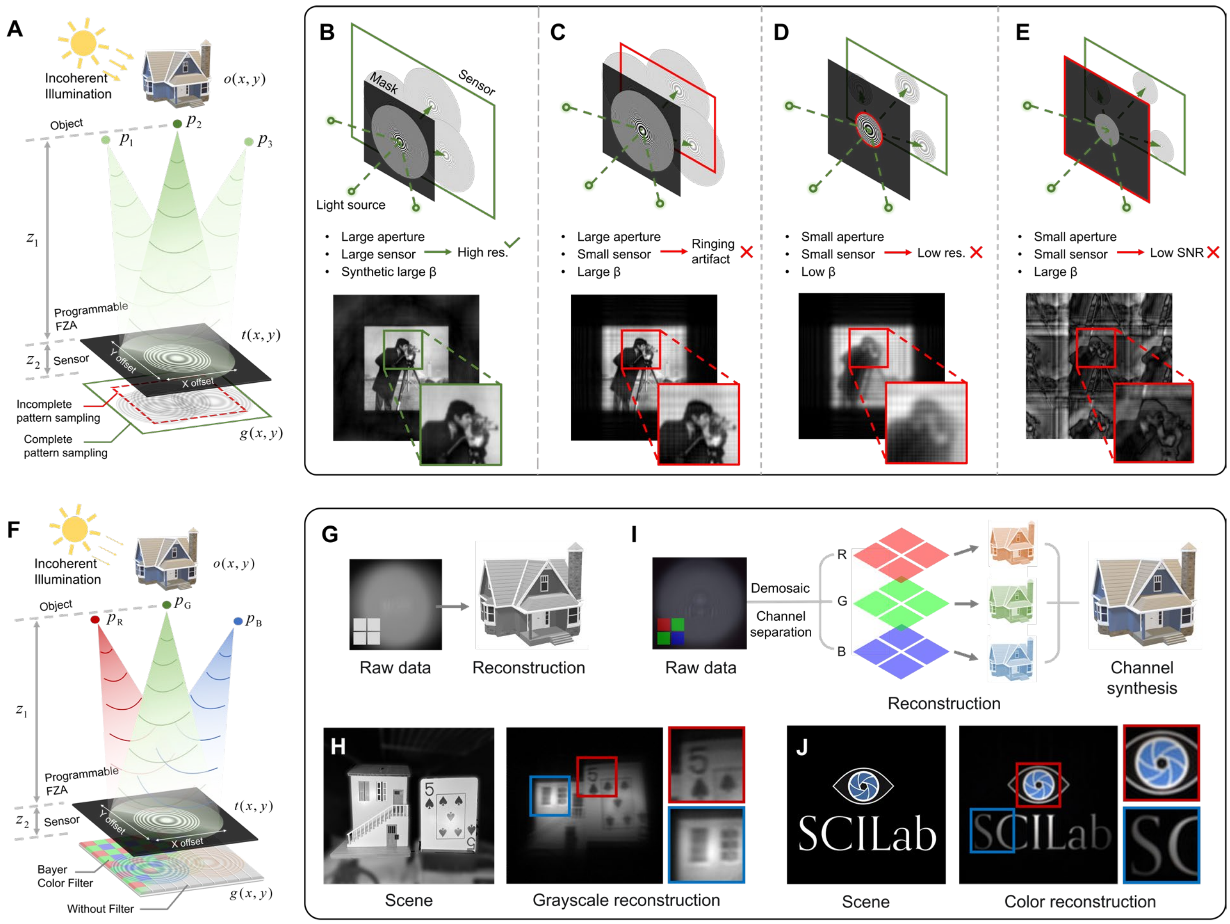

Fig. S7 Considerations for raw data during the measurement process.

(**A**) The impact of sensor measurement overflow on measurement and reconstruction. (**B**) Reconstruction results with all parameters optimally matched. (**C**) Reconstruction result with incomplete pattern sampling by the sensor. (**D**) Reconstruction results with an appropriate aperture size and a smaller β parameter. (**E**) Reconstruction results with an appropriate aperture size and a larger β parameter. (**F**) The impact of color and spatial information characteristics on measurement and reconstruction. (**G-H**) Raw data captured and reconstructed with a grayscale sensor. (**I-J**) Raw data captured with a color sensor and reconstructed by channel separation.

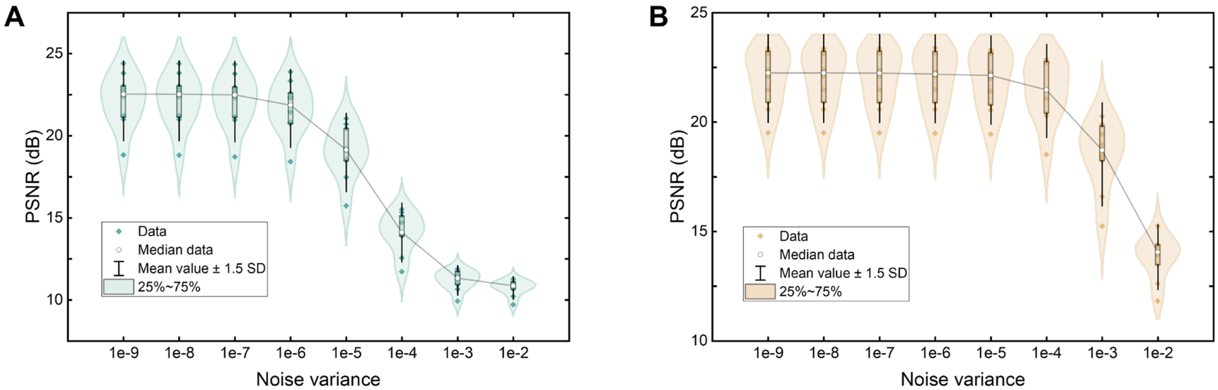

Fig. S8 Quantitative comparison of LIP in (A) dynamic mode and (B) static mode under noise variance σ = 1e-9 ~ 1e-2.

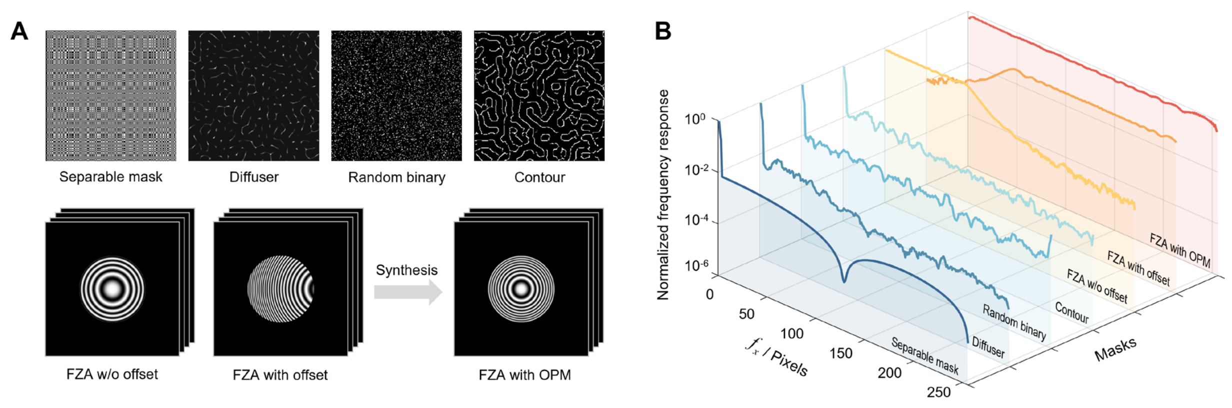

Fig. S9 Comparison of the FZA-based lensless imaging method and traditional static-mask lensless imaging methods (separable mask, diffuser, random binary, and contour) regarding PSF pattern and frequency response.

(**A**) The mask patterns corresponding to different masks, where the FZA-based methods include four distinct PSF patterns for four-step phase shifting, and the OPM-based FZA method achieves synthesis by combining sub-apertures with different offsets. (**B**) The normalized logarithmic frequency response curves for each mask, with all mask sizes standardized to 512×512 pixels. Due to the symmetry of the frequency spectrum, only the positive fx axis range [0, 256] is displayed for comparison.

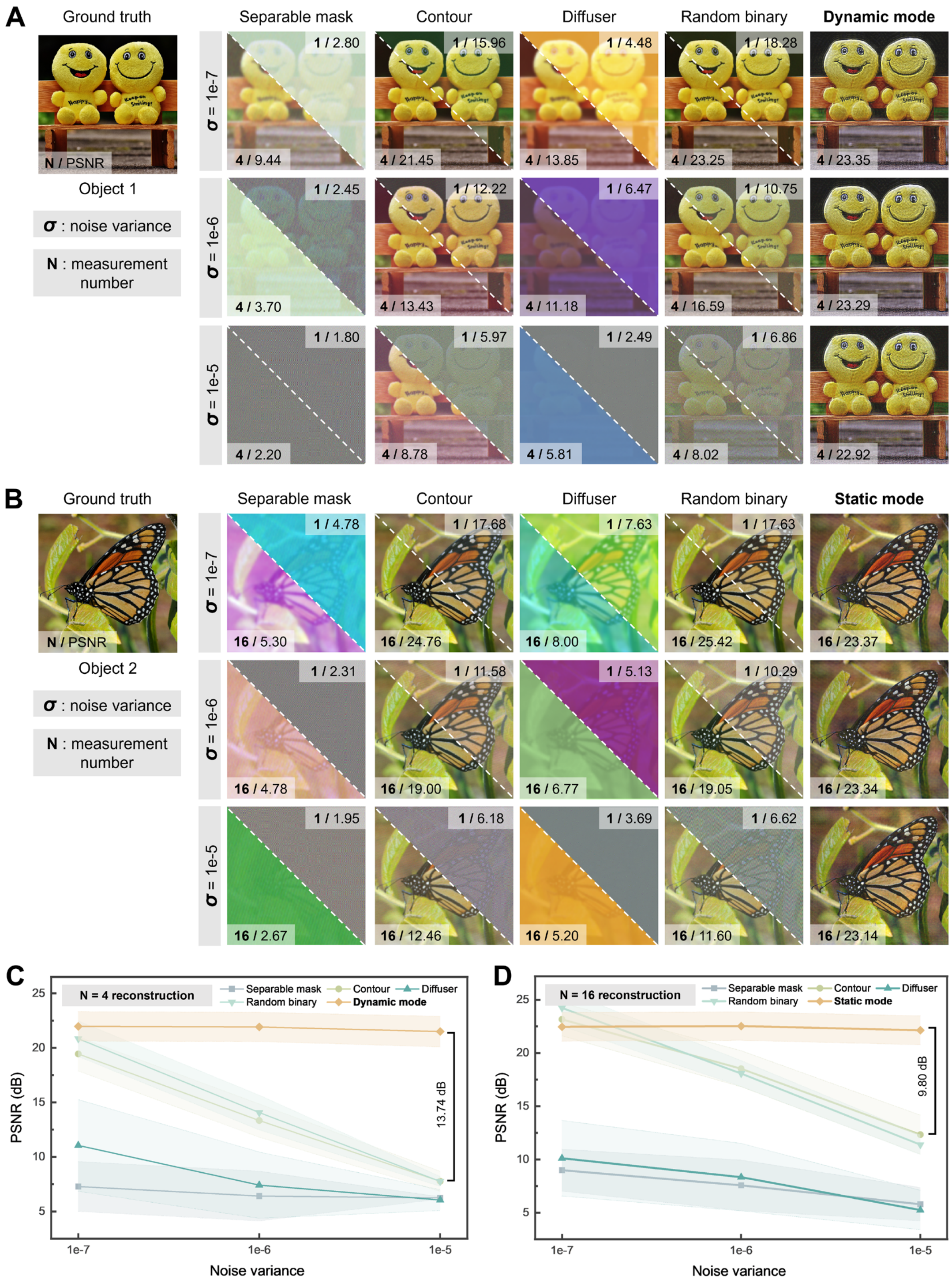

Fig. S10 Comparison of the proposed method and traditional static-mask lensless imaging methods (separable mask, contour, diffuser, and random binary) in numerical simulations of complex colored scenes.

(**A**) Comparison of reconstruction results of Object 1 between the dynamic mode (N = 4) and traditional static-mask lensless imaging (N = 1 / N = 4), with added zero-mean Gaussian noise under variance σ = 1e-7, 1e-6, and 1e-5. (**B**) Comparison of reconstruction results of Object 2 between the static mode (N = 16) and traditional static-mask lensless imaging (N = 1 / N = 16), with added zero-mean Gaussian noise under variance σ = 1e-7, 1e-6, and 1e-5. (**C**) PSNR comparison curves of the dynamic mode (N = 4) and traditional static-mask lensless imaging with a 4-frame average under different noise variances (N = 4). (**D**) PSNR comparison curves of the static mode (N = 16) and traditional static-mask lensless imaging with a 16-frame average under different noise variances (N = 16). Photo credit: PxHere, Pixnio.

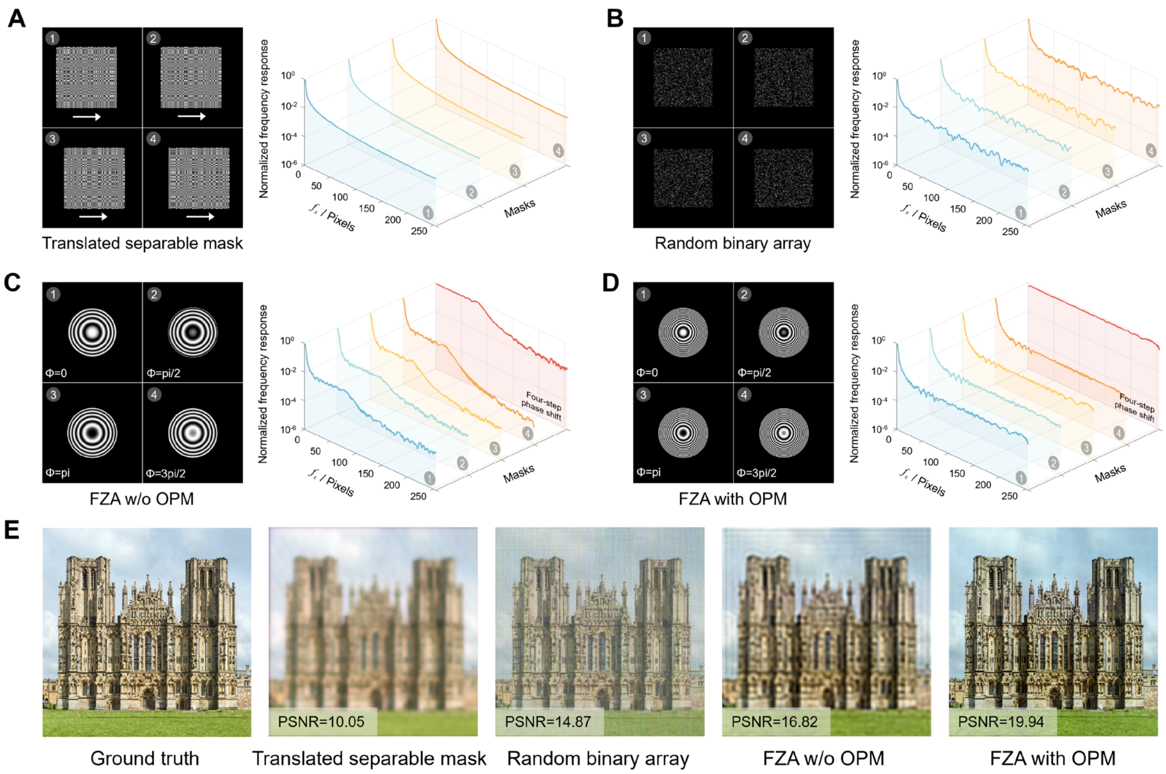

Fig. S11 Comparison of the FZA-based lensless imaging method and multi-shot lensless imaging methods (translated separable mask and random binary array) regarding mask, frequency response, and reconstruction.

The mask patterns and corresponding frequency response of (**A**) translated separable mask, (**B**) random binary array, (**C**) FZA without OPM, and (**D**) FZA with OPM, where the FZA-based methods include four distinct masks for four-step phase shifting, and the OPM-based FZA method achieves synthesis by combining sub-apertures with different offsets. All mask sizes are standardized to 512×512 pixels. Due to the symmetry of the frequency spectrum, only the positive fx axis range [0, 256] is displayed for comparison. (**E**) Reconstruction results of different multi-shot methods, each method carried out 16 measurements, and added zero-mean Gaussian noise with variance of 1e-9 to each measurement to simulate the real measurement scene. Photo credit: DAVID ILIFF.

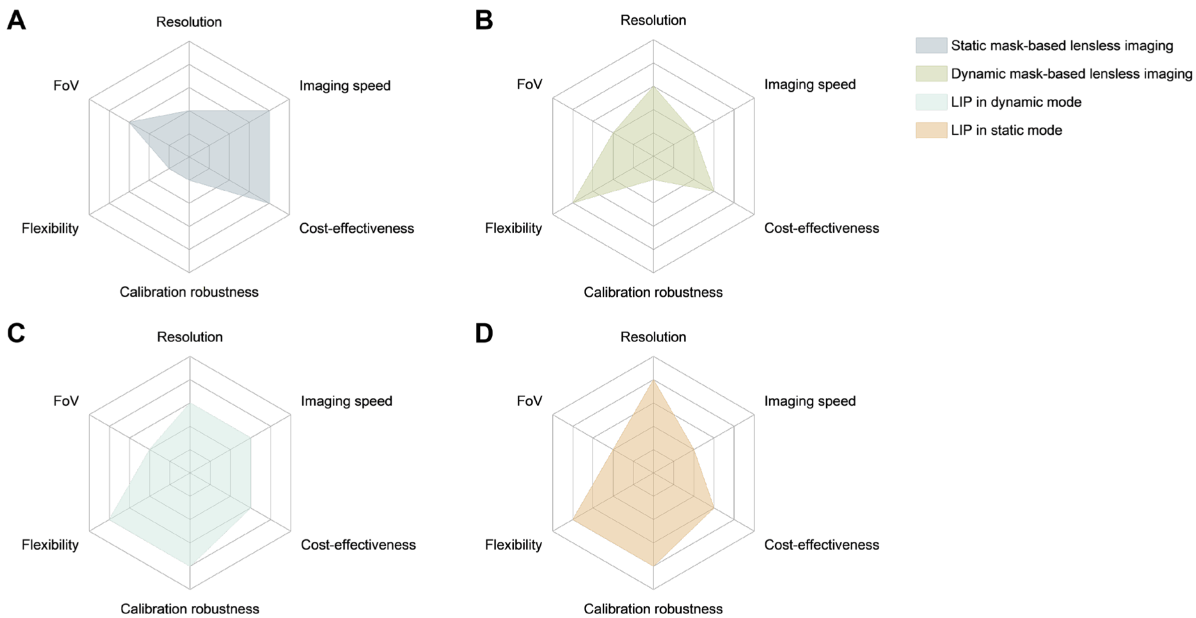

Fig. S12 Qualitative comparison of (A) static mask-based lensless imaging, (B) dynamic mask-based lensless imaging, (C) LIP in dynamic mode, and (D) LIP in static mode across six aspects, including "Resolution," "Imaging speed," "FoV," "Cost-effectiveness," "Calibration robustness," and "Flexibility."

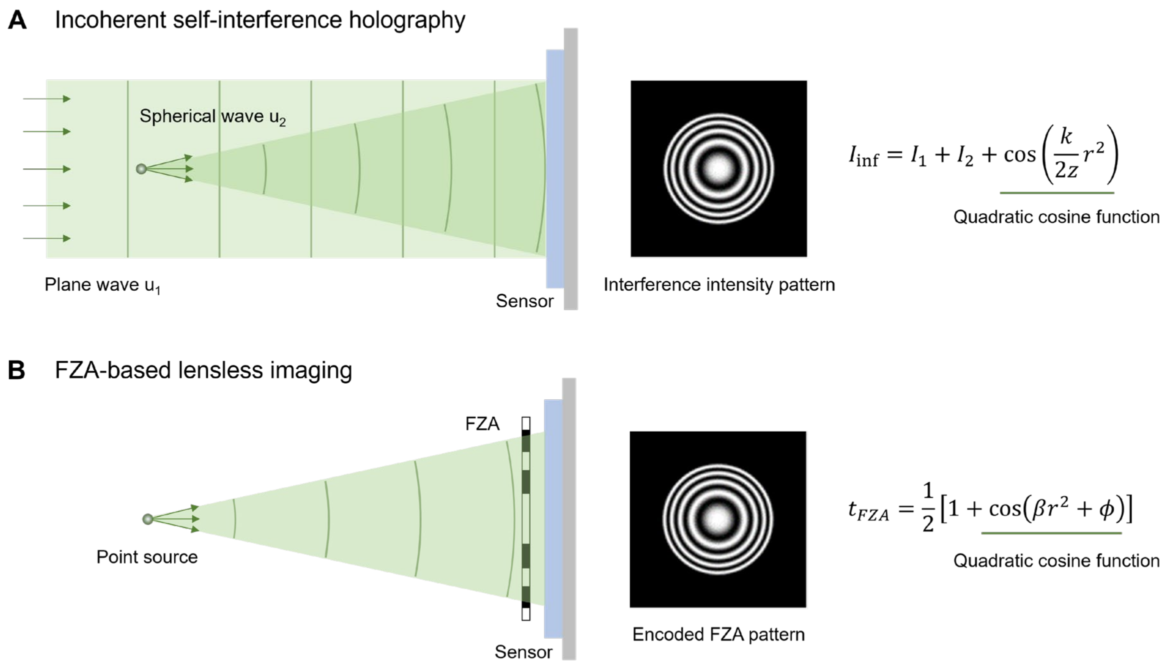

Fig. S13 The connection between incoherent self-interference holography and FZA-based lensless imaging.

(**A**) The principle of incoherent self-interference holography: The interference intensity pattern is derived from the self-interference between the plane wave and the spherical wave of the object. (**B**) The principle of FZA-based lensless imaging: The encoded pattern is formed by the projection of a point source through the FZA.

Movie S1.
Forward model of programmable FZA-based lensless imaging system.

Movie S2.
Optimal parameter matching and offset-FZA parallel merging for lensless imaging with programmable FZA.

Movie S3.
Dynamic scene experiment of LIP.

Movie S4.
Hand gesture interaction with LIP module.