


Differential ghost imaging with learned modulation patterns

Xiao Wang¹, Pengxiang Cheng, Huaijian Chen, Shupeng Zhao, Guangdong Ma, Yongchang Zhang¹,
Pei Zhang¹, Hong Gao, Ruifeng Liu^{1,*} and Fuli Li

*Ministry of Education Key Laboratory for Nonequilibrium Synthesis and Modulation of Condensed Matter,
Shaanxi Province Key Laboratory of Quantum Information and Quantum Optoelectronic Devices, School of
Physics, Xi'an Jiaotong University, Xi'an 710049, China*

 (Received 19 January 2024; revised 8 April 2024; accepted 6 June 2024; published 10 July 2024)

Unlike conventional imaging with two-dimensional array sensors featuring millions of pixels, ghost imaging enables the use of advanced detector technologies, giving advantages such as high signal-to-noise ratio, wide spectral range, and robustness to light scattering. However, this involves an extremely time-consuming measurement process, which means that it is difficult to meet the needs of high-quality real-time imaging. This paradox becomes notable especially in the context of utilizing non-orthogonal modulation patterns, such as the speckles generated by rotating ground glass. Efficient modulation patterns and advanced reconstruction algorithms are widely studied as two main ideas to solve the above problem. Here, we perform real-time, high-fidelity differential ghost imaging (DGI) at a low sampling ratio of 6.25% by proposing a compact physically guided single-layer neural network with the DGI algorithm embedded. Simulations and experiments show that, once the learned modulation patterns are obtained, our scheme can achieve fast, high-quality, and noise-robust DGI without the need for complex iterative optimization algorithms or subsequent optimization neural networks. Our scheme opens up new horizons for exploring more efficient modulation patterns for ghost imaging by deeply combining physical priors.

DOI: [10.1103/PhysRevApplied.22.014023](https://doi.org/10.1103/PhysRevApplied.22.014023)

I. INTRODUCTION

Ghost imaging (GI) restores a scene by calculating the second-order correlation of two light beams: modulation pattern and bucket signal [1–15]. GI was first demonstrated by Pittman *et al.* using entangled photon pairs [1]. Subsequently, Bennink *et al.* discovered that GI could also be achieved using classical light sources [2]. This discovery sparked profound investigations into the nature of GI within the academic community [3–6]. The eventual realization of computational ghost imaging (CGI) [7,10] demonstrates that GI can be accurately described using second-order correlations of classical optical fields, without the necessity of relying on quantum properties.

Additionally, Shapiro's CGI [7] illustrates that GI can be achieved using only a single-pixel detector through pre-computation of modulation patterns. It is worth noting that, earlier, Sen *et al.* proposed a dual-photography [16] image acquisition method that employed a single-pixel detector to measure the intensity of modulated light. This approach can be considered a prototype of single-pixel imaging [8,12,13,17], closely resembling GI. Since there is only one pixel to collect all photons, GI with a single-pixel detector is endowed with great advantages over traditional camera imaging, such as detection sensitivity, dark

count, response time, and spectral range. These advantages are valuable in low-detection-intensity situations such as medical imaging, remote sensing, or nonvisible spectral scenarios where array detectors are expensive and rare.

Naturally, GI has received extensive attention from researchers in various fields, such as spectral imaging [18–20], terahertz imaging [17,21–25], x-ray imaging [26–28], remote sensing [29,30], object tracking [31–33], three-dimensional (3D) imaging [19,34,35], complex amplitude optical field reconstruction [36–39], imaging through scattering media [40–42], photoacoustic imaging [43–45], etc. But that which makes also breaks: it is also because only one pixel collects the light intensity of the object, the reconstruction of the object requires multiple measurements. This leads to a non-negligible acquisition time, and one has to make a trade-off between imaging quality and measurement time, hindering the application of GI in practical scenarios.

The need for fast and high-quality GI has led to several outstanding efforts to address the above issues. On the one hand, methods like speckle patterns with different speckle sizes [46], reordering of the Hadamard bases, [47–50] or Fourier bases of downsampled Fourier spectrum [51] are committed to designing and using more efficient modulation patterns to achieve high-quality imaging with a small number of measurements. On the other hand, methods such as alternating projection [52] and

*Contact author: ruifeng.liu@mail.xjtu.edu.cn

compressed sensing [53,54] based techniques aim to obtain better reconstruction by developing more advanced optimization algorithms. However, this usually leads to multiple iterations and high computational complexity.

Recently, deep learning (DL) [55] in an end-to-end form has been adopted to construct the mapping from low-quality reconstruction [56–58] or bucket signal [59,60] to high-quality output. Furthermore, some researchers have added modulation patterns into network training to conduct imaging or sensing with the trained modulation patterns [61–67]. Typically, a subsequent optimization network is employed to intricately establish a mapping that bridges the bucket signals [63] or lower-quality reconstructions [66] with high-fidelity reconstructions.

In early GI studies, images were directly reconstructed with noniterative intensity correlation algorithm [1,4,7,10,11]. For instance, in CGI [7], one considers a two-dimensional image of an object, $O \in \mathbb{R}^{L \times L}$, consisting of N pixels, where $N = L \times L$. A sequence of modulation patterns $H = [H_1, H_2, \dots, H_M] \in \mathbb{R}^{M \times L \times L}$ are loaded onto a spatial light modulator. Here $H_m \in \mathbb{R}^{L \times L}$ represents the m th modulation pattern, and M denotes the total number of modulation patterns. After interacting with the object, M bucket detection signals are obtained, denoted as $I = [I_1, I_2, \dots, I_M] \in \mathbb{R}^M$. The imaging result can be obtained through the following correlation operation: $O = \langle H_m I_m \rangle - \langle I_m \rangle \langle H_m \rangle$, where $\langle \cdot \rangle$ is the ensemble average in terms of m . Here, the reconstruction of objects is the weighted sum of modulation patterns, with the bucket detection signals serving as weights. A larger measurement signal represents a greater similarity between the modulation pattern and the target scene, thus carrying more weight in the reconstruction. These algorithms strictly fit the physical mechanisms of GI and their noniterative nature leads to almost negligible reconstruction times, making them well suited to meet the demands of real-time imaging.

It is worth noting that, as the differential measurement is sensitive only to the fluctuating part of the intensity signal, differential ghost imaging (DGI) [11] can significantly enhance the signal-to-noise ratio (SNR) in contrast to conventional GI. In contrast to the aforementioned DL methods, the DGI algorithm is simpler and is unencumbered by the well-documented issue of interpretability associated with neural networks. Additionally, the algorithm possesses inherent flexibility. The sequence of modulation pattern projections need not be fixed, and real-time reconstruction can be achieved using only a limited number of modulation patterns, eliminating the need to wait for the completion of all modulation pattern projections. Furthermore, algorithms such as DGI are devoid of any nonlinear operations, a quality that can be advantageous in the context of optical computations, where the incorporation of nonlinearities presents a formidable challenge. While DGI offers numerous advantages, achieving high-quality DGI with a low sampling ratio requires more

efficient modulation patterns and is still an urgent issue to be addressed.

Is there a set of superior modulation patterns available that can enhance the quality of DGI at a low sampling ratio without augmenting the algorithm's intricacy? To address this question, we propose a compact physics-guided single-layer neural network for DGI with learned grayscale modulation patterns at a low sampling ratio of 6.25%. The design of our network structure completely follows the two processes of DGI: bucket signal acquisition and DGI reconstruction. In addition, we add physical prior regularization to the network training, which effectively alleviates the artifacts and noise in the reconstructed image and improves the noise robustness of the learned modulation patterns. Benefiting from rigorous physical modeling and the concise network structure, our learned modulation patterns can be directly used for real-time high-quality DGI without the need for additional optimization networks or complex iterative optimization algorithms. Compared with the typical modulation patterns, our scheme shows superiority in terms of imaging quality and noise robustness at a low sampling ratio of 6.25%.

II. METHODS

The schematic diagram of the experimental setup is shown in Fig. 1(b). The light emitted by the light source (a white light-emitting diode lamp) is reflected by the object and then imaged to a digital micromirror device (DMD; DLP7000, 1024×768 pixel array of $13.6 \times 13.6 \mu\text{m}^2$ mirrors) by an imaging optic with a focal length of 100 mm. After being modulated by the DMD loaded with the modulation patterns, the intensity signals are sequentially collected by a single-pixel detector (Thorlabs PDA100A-EC) with a focusing lens of 50 mm.

The bucket signal acquisition can be expressed as

$$I^m = \sum_{i,j}^L H_{ij}^m \cdot O_{i,j} \quad (1)$$

where I^m and H^m are the m th bucket signal and modulation pattern, respectively, and O is the object. Here we assume that the size of the object is $L \times L = N$, with a total of N pixels. We adopt M modulation patterns; then the sampling ratio can be defined as $\beta = M/N$. In all our simulations and experiments, $L = 128$, $N = 16384$, and $M = 1024$, so we have $\beta = 6.25\%$. By vectorizing O and I , we can get the matrix representation of Eq. (1):

$$\mathbf{I}^{M \times 1} = \mathbf{H}^{M \times N} \mathbf{O}^{N \times 1}. \quad (2)$$

Here $\mathbf{I}^{M \times 1}$ is the column vector composed of M bucket signals, $\mathbf{H}^{M \times N}$ is the modulation matrix composed of M modulation patterns, and $\mathbf{O}^{N \times 1}$ is the column vector composed of N pixels of the object. One common and intuitive

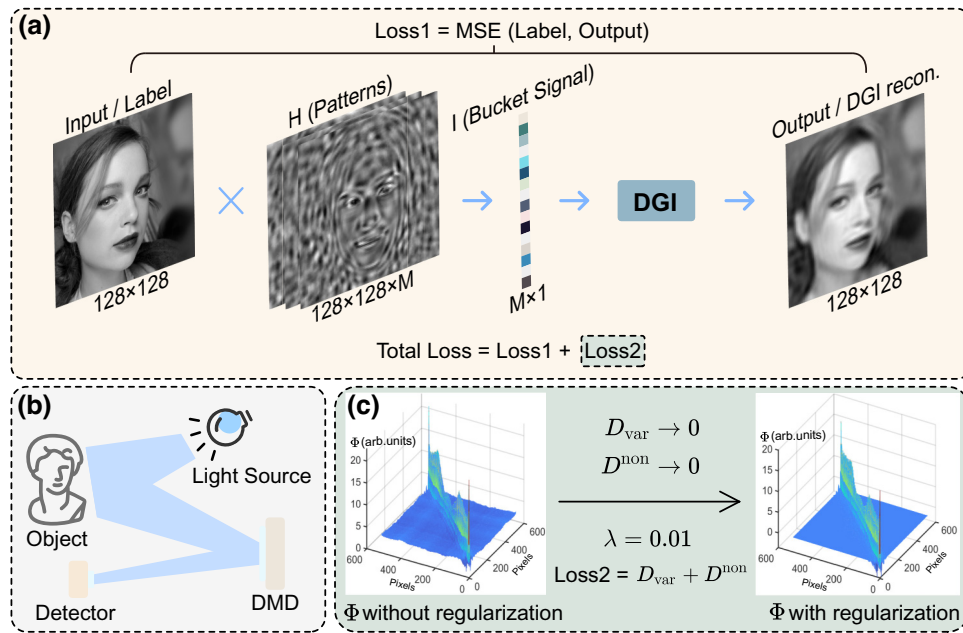


FIG. 1. Principle and framework diagram of the proposed scheme. (a) The compact physics-guided single-layer neural network with learned grayscale modulation patterns. DGI: differential ghost imaging. (b) A brief sketch of the experimental setup. (c) Φ regularization is added to the optimization of the objective function (loss) as an *a priori* constraint. Specifically, we mitigate the artifacts and noise in the reconstruction by constraining the variance of the diagonal elements and nondiagonal elements of the matrix Φ to converge to zero.

method is using an inverse matrix \mathbf{H}^{-1} to reverse the above linear model. However, $M < N$ in our case, which means the modulation matrix \mathbf{H} is nonsymmetric. To make this problem solvable, we multiply by $(\mathbf{H}^{M \times N})^T$ on both sides of Eq. (2) and let $\Phi = (\mathbf{H}^{M \times N})^T \mathbf{H}^{M \times N}$, so we get

$$(\mathbf{H}^{M \times N})^T \mathbf{I}^{M \times 1} = \Phi^{N \times N} \mathbf{O}^{N \times 1}. \quad (3)$$

It is not difficult to see from Eq. (3) that a good solution can be obtained when Φ approaches the identity matrix. However, $M < N$ in our case, as mentioned above, so it is theoretically impossible for matrix Φ to reach the ideal identity matrix; that is, the matrix Φ can only approximate the identity matrix as closely as possible with a theoretical limit.

In our scheme, we split the matrix Φ into the sum of two matrices (Φ_{diag} and Φ_{nondiag}) with diagonal and nondiagonal elements of matrix Φ , respectively, as shown in Fig. 2. We argue that, for a Φ that infinitely closely approaches an identity matrix, the Φ_{diag} matrix with flat diagonal elements contributes to good reconstruction, while the Φ_{nondiag} matrix with nonzero elements mainly contributes to artifacts and noise in the reconstructed image. Based on this, we require that the Φ matrix should approach the identity matrix by restricting the variance of the diagonal elements, and the absolute sum of the nondiagonal

elements of Φ should go to zero, as shown in Fig. 1(c):

$$\begin{aligned} D_{\text{var}} &= \text{diag}(\Phi)_{\text{var}} \rightarrow 0, \\ D_{\text{non}} &= \frac{1}{N(N-1)} \sum |\text{diag}^{\text{non}}(\Phi)| \rightarrow 0. \end{aligned} \quad (4)$$

Here, we design a single-layer neural network in strict accordance with the physical process of DGI, which contains two parts: bucket signal acquisition and DGI reconstruction, as shown in Fig. 1(a). In the first part, the object \mathbf{O} serves as the network input and label simultaneously, and is modulated by M modulation patterns \mathbf{H} to produce M bucket signals \mathbf{I} : $\mathbf{I} = \mathbf{H}\mathbf{O}$. In the second part, we choose the DGI, a noise robust algorithm, combined with the modulation patterns and bucket signals from the first part to reconstruct the object image. Only the modulation patterns in the first part of the network are trained as learnable parameters, so it is considered as a single-layer network. The loss function of the network consists of two parts. In addition to the commonly used mean square error (MSE) between the outputs and the labels, we innovatively introduce the matrix Φ regularization to optimize the modulation patterns as shown in Fig. 1(c). We alleviate the artifacts and noise in the reconstructions by constraining the variance of diagonal elements and nondiagonal elements of Φ to approach zero, so as to improve the imaging quality of the learned modulation patterns.

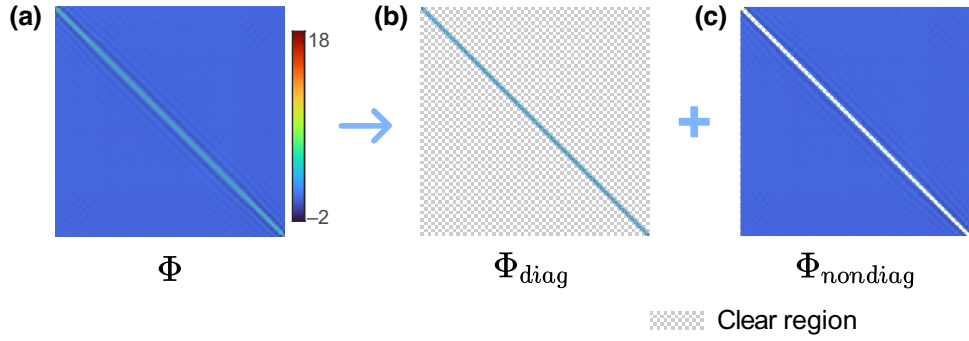


FIG. 2. Schematic diagram of the Φ splitting. (a) Representation of the matrix Φ . In (b),(c), the blue sections correspond to the diagonal and nondiagonal elements of the split Φ matrix, respectively. To form the matrices Φ_{diag} and Φ_{nondiag} , the clear regions in (b),(c) are filled with zeros.

Therefore, the final objective function (loss function) to optimize the modulation patterns H is

$$H^* = \operatorname{argmin}_H \|\text{DGI}(H, HO) - O\|^2 + \lambda(D_{\text{var}} + D^{\text{non}}). \quad (5)$$

Here H^* is the optimized modulation pattern, $\text{DGI}(\cdot)$ represents the DGI reconstruction algorithm, and λ is the weight factor used to balance Φ regularization in the objective function. We choose $\lambda = 0.01$ in our scheme, and more discussion about λ can be found in the Appendix C.

We train the network on two datasets, CelebAMask-HQ [68] and STL-10 [69]. CelebAMask-HQ is a dataset consisting of 30 000 high-resolution human faces, 29 000 of which are used as the training set and the remaining 1000 are used as the test set after being downsampled to 128×128 . STL-10 is a dataset containing 100 000 natural images of 10 categories with a size of 96×96 . After upsampling to 128×128 , 90 000 (10 000) of them are used as the training set (test set). The specific training process is as follows. Firstly, the image samples (also serving as labels) are input into the network. Subsequently, the trainable parameters in the network, i.e., the modulation patterns H , are optimized via the objective function [Eq. (5)]. Finally, upon completion of training, the network parameters H^* are fixed and extracted, combined with the DGI reconstruction algorithm for imaging. The whole training process is implemented on the Pytorch framework and sped up by an NVIDIA RTX 3090 graphics processing unit. Our code is available at [70].

III. RESULTS

It should be emphasized that the core of our scheme is to explore more efficient modulation patterns using DGI methodology at a low sampling ratio of 6.25%. We focus on comparing the grayscale modulation patterns obtained by our scheme with several typical modulation patterns in GI: binary random patterns (Random), reordered

Hadamard patterns (Hadamard) [48], Fourier sinusoidal patterns (Fourier) [51], the trained binary patterns of the deep convolutional autoencoder network (DCAN) [63], and the physics-informed deep neural network (Informed and Fine-tune) [66]. Specifically, our scheme obtains four sets of modulation patterns by training on two datasets: human face dataset CelebAMask-HQ [68] and natural images of 10 categories dataset STL-10 [69]. We mark the four sets of modulation patterns by Face, Face-Reg, STL10, and STL10-Reg, where “-Reg” indicates that Φ regularization is added during the training process. Figure 3 row 7 shows the examples of different modulation patterns. See Appendix B for more details on the learned modulation patterns. We perform numerical simulations with these modulation patterns under the same sampling ratio of $\beta = 6.25\%$, and the reconstruction results for several objects are shown in Fig. 3.

Obviously, as shown in Fig. 3 column a, it is difficult to extract any target information from the reconstruction of the random binary patterns (Random) with such a low sampling ratio. Because of the low sampling ratio, the Hadamard reconstructions show mosaic-like results with a lot of details lost, as shown in Fig. 3 column b. The Fourier reconstruction [Fig. 3 column c] also exhibits loss of high-frequency details accompanied by severe ringing effects [71]. The reconstruction of DCAN and Informed (trained with the dataset STL10), as shown in Fig. 3 columns d and e, respectively, have similar performance. They exhibit a slight improvement over Hadamard but fall short of Fourier’s reconstruction. In addition, such schemes using subsequent optimization networks have the possibility of hallucinations and creation of artifacts [72]. These artifacts are known to be features that do not look real, which is evident in the results of Informed shown in Fig. 3 column e. Yet, as shown in Fig. 3 column f, the reconstruction quality was significantly improved and the above artifacts were alleviated after adopting the iterative optimization Fine-tune process based on Informed. As can be seen from Fig. 3 rows 1 and 4 and their local

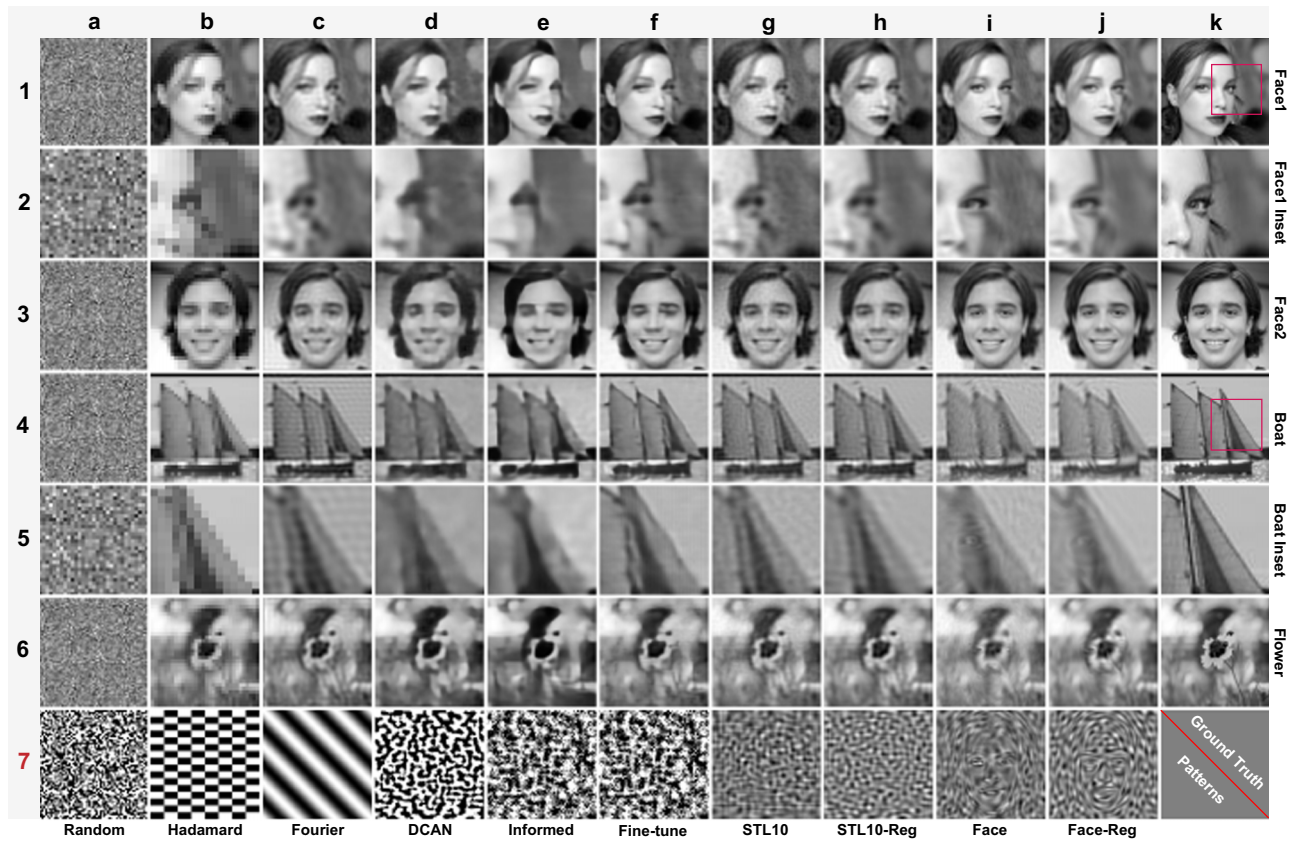


FIG. 3. Comparison of simulated reconstruction results of different modulation patterns with a sampling ratio of $\beta = 6.25\%$. Row 7 contains the examples of different modulation patterns, and column k shows the ground truth for objects. Rows 1 and 3 are face targets, while rows 4 and 6 are nonface natural images. Rows 2 and 5 are partial magnifications of rows 1 and 4, respectively, and the red boxes in k1 and k4 are the selected magnification areas.

enlargements in Fig. 3 rows 2 and 5, our schemes show fewer artifacts and distortions while having better details compared to Hadamard, Fourier, DCAN, and Informed results, especially Face-Reg and STL10-Reg with Φ regularization. The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) for Face-Reg and STL10-Reg with Φ regularization also show an overall improvement compared to Face and STL10 shown in Figs. 4(a) and 4(b).

Specifically, Face and Face-Reg trained with the human face dataset perform well on face images (Face1 and Face2), as shown in Fig. 3 column i, rows 1–3 and column j, rows 1–3, which are rich in facial details, smooth, and detailed overall. However, this advantage does not continue in nonface images. Because of the intrinsic characteristics of the modulated patterns shown in Fig. 3 column i, row 7 and column j, row 7, the nonface reconstructions [Fig. 3 column i, rows 4–6 and column j, rows 4–6] of Face and Face-Reg appear with artifacts of facial-like features. This sacrifice of generalizability in exchange for outstanding performance for specific tasks has certain implications in scenarios such as magnetic resonance

imaging. In contrast, STL10 and STL10-Reg trained with the STL-10 dataset have balanced and decent performance on face and nonface images [Fig. 3 columns g and h], but are slightly inferior to Face and Face-Reg on face images. This is also consistent with the quantitative results in Figs. 4(a) and 4(b).

The above results are obtained without considering noise, and noise is an important factor that cannot be ignored in practical application scenarios. Here, we simulate the noise conditions by adding Gaussian white noise to the bucket signal and use the SNR (where $\text{SNR} = 10 \log_{10}(P_{\text{signal}}/P_{\text{noise}})$, where P_{signal} and P_{noise} are the powers of the bucket signal and noise, respectively) of the bucket signal to measure the noise level. We selected a total of 20 images containing both human faces and natural images as test data. The average PSNR and SSIM curves of Figs. 4(c) and 4(d) show that the imaging quality of all modulation patterns deteriorates as the SNR decreases (i.e., the noise increases). At a very low sampling ratio, Random cannot achieve effective reconstruction; consequently, its results exhibit minimal variation under different noise conditions. Fourier is very sensitive to noise,

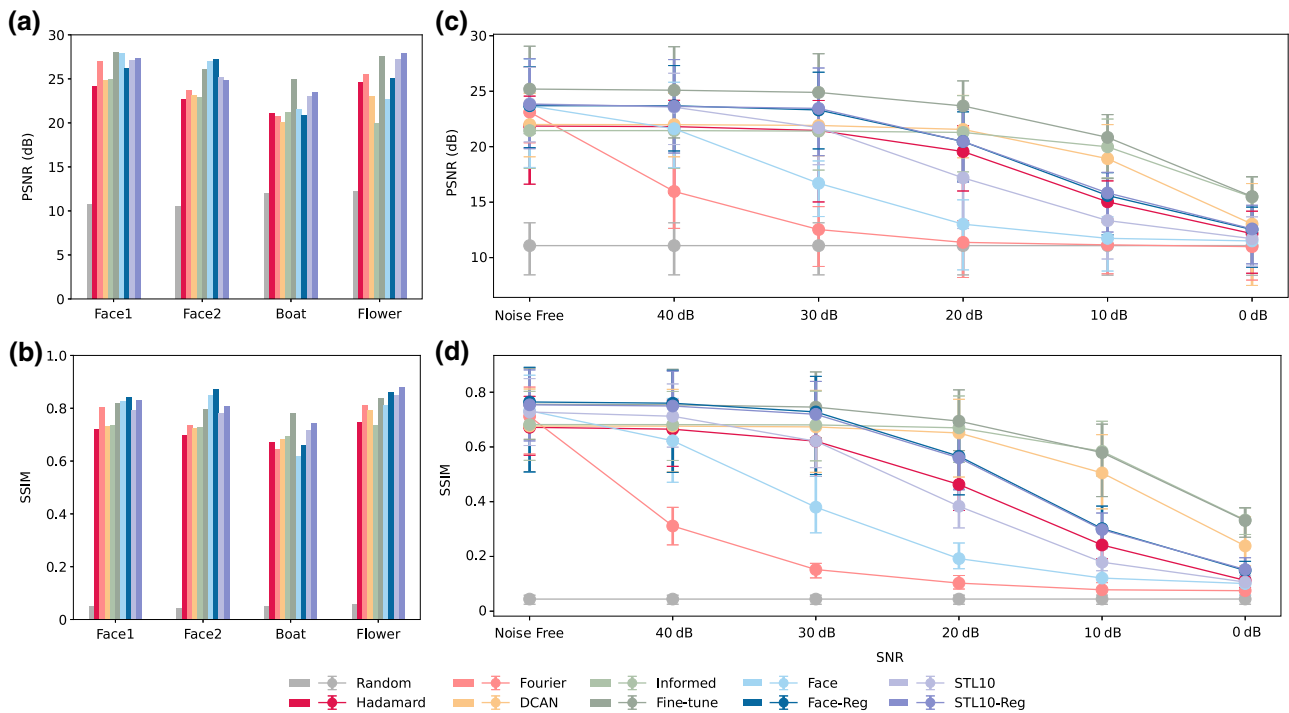


FIG. 4. Quantitative comparison of imaging quality for different modulation patterns with $\beta = 6.25\%$. (a),(b) The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) for the reconstructions of target images (Face1, Face2, Boat, and Flower). (c),(d) The curves of average PSNR and SSIM versus noise for the reconstructions of 20 selected face and natural target images.

and its reconstruction quality decreases greatly with the increase of noise.

In contrast, three schemes based on subsequent optimization network, DCAN, Informed and Fine-tune, show strong robustness to noise. This is not surprising, since the strong noise immunity of neural networks is a well-known fact. Compared with the above three network schemes,

our scheme performs worse, especially under strong noise conditions. However, our scheme has considerable advantages over other schemes (except the Fine-tune) at low noise conditions. Specifically, the reconstruction of Face and STL10 without Φ regularization degrades significantly with the enhancement of noise, while Face-Reg and STL10-Reg show comparable noise robustness to

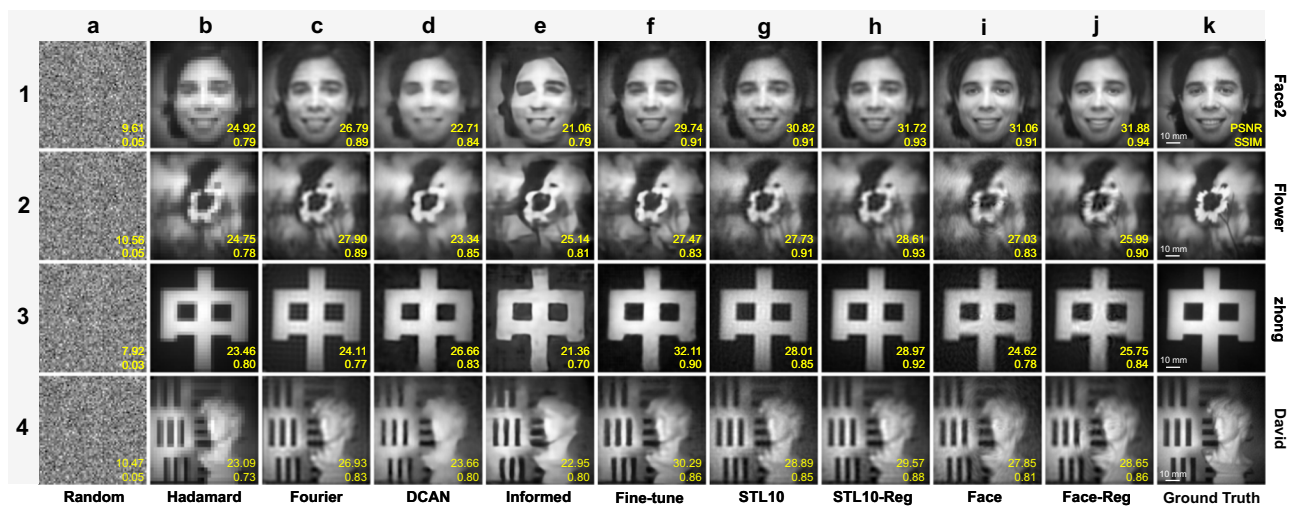


FIG. 5. Comparison of the experimental reconstruction results of different modulation patterns with a sampling ratio of $\beta = 6.25\%$. Face2, Flower, and the Chinese character “zhong” are targets printed on A4 paper, and David is a 3D sculpture with part of the USAF resolution chart as the background. The lower right corner of the reconstructed images shows PSNR and SSIM values, respectively.



FIG. 6. A selection of image frames of the 30-s video (see Visualization 1 in the Supplemental Material [73]) with the STL10-Reg modulation patterns. The lower right green number is the corresponding frame number.

Hadamard with a slight advantage under Φ regularization. It can be seen that Φ regularization in our scheme can effectively improve the noise robustness of the modulation patterns. In fact, on average, Fine-tune has the best performance of all the schemes as shown in the quantitative indices in Figs. 4(c) and 4(d). However, this iterative approach comes at the expense of time, making it difficult to meet the requirements of real-time imaging.

In experiments, we adopt the experimental system shown in Fig. 1(b). Like the settings in simulation, we compare the actual imaging performance of several schemes at a sampling ratio of 6.25%. Details on the preprocessing for modulation patterns in experiment and reconstruction algorithms can be found in the Appendix A. Figure 5 shows the experimental reconstruction for different modulation patterns. Here we select two types of objects as imaging targets: one is a target printed on A4 paper, such as Face2, Flower, and the Chinese character “zhong”; the other is a real 3D object, the sculpture David with part of the USAF resolution chart as the background. We use the image reconstructed by the Hadamard bases with a sampling ratio of 100% as the ground truth shown in the last column of Fig. 5. The quantitative evaluation PSNR and SSIM of the reconstructed images are marked in the lower right corner.

Overall, Fig. 5 shows the results consistent with the simulation shown in Fig. 3. After time-consuming iterative optimization, Fine-tune has been greatly improved compared with Informed. Our scheme still achieves the best reconstruction quality except for Fine-tune. Face and Face-Reg trained with the human face dataset yield more detailed facial features when they are used to reconstruct Face2, but perform slightly worse on nonface images. STL10 and STL10-Reg trained with the natural image dataset have decent and balanced performance on various categories of images. By comparing the reconstruction results of STL10 and STL10-Reg, as well as Face and Face-Reg, we can draw the same conclusion as the simulation: after adding Φ regularization, the quality of

reconstructed image is improved, as the image is smoother and artifacts and noise are effectively suppressed.

Since the proposed scheme uses the learned modulation patterns for DGI, its imaging speed is mainly limited by the refresh rate of the spatial light modulator. Here, we experimentally verify the real-time imaging performance of the proposed scheme with STL10-Reg modulation patterns. We choose a toy astronaut as imaging target and capture its dynamic motion with part of the USAF resolution chart as the background. We set the DMD to operate at its highest refresh rate: 22 kHz. For images with a pixel number of $N = 128 \times 128$, we achieve real-time imaging with a video rate of ~ 10 Hz at a sampling ratio $\beta = 6.25\%$. We captured a 30-s-long video with 300 frames, and 10 example frames are shown in Fig. 6. The results in Fig. 6 show that the overall video imaging results are good, with clear details. However, the image also appears noisy (e.g., frames 205, 225, and 265) when the target object is moving fast. The full video is available in Visualization 1 in the Supplemental Material [73].

IV. DISCUSSION AND CONCLUSION

We recall two aspects of the efforts in GI: better modulation patterns and more advanced reconstruction algorithms. Compared with methods committed to better modulation patterns [46–51], our scheme attains superior image quality and noise robustness while maintaining a comparable simple reconstruction algorithm as these methods do. As for schemes dedicated to the development of more advanced reconfiguration algorithms [52–54], this usually leads to multiple iterations and high computational complexity. Furthermore, existing schemes [63–66] for training modulation patterns are limited to binarization constraint on the modulation patterns, while additional subsequent optimization networks are required (e.g., Unet). The binary modulation pattern has only two values for its elements, and its variation space is greatly restricted. This leads to a limited encoding capacity for

modulation patterns, especially at a low sampling ratio. Alternatively, the Fine-tune approach [66] enhances the Informed [66] by iteratively optimizing the reconstruction quality for an individual target image. Nevertheless, such excellent image quality is accompanied by a drawback of an iterative process that consumes considerable time, which poses a challenge to cater to the requirements of real-time imaging.

Compared to these methods, our approach offers a notably simpler and more interpretable reconstruction process, yet still achieves optimal imaging quality (except for Fine-tune). Our model stands out due to its fewer training parameters, resulting in a faster training and reconstruction speed. Furthermore, the DGI of the reconstruction part does not contain any nonlinear operations, which may be of benefit in optical computations where nonlinearities are difficult to introduce. We also acknowledge that, in scenarios with high noise levels, the imaging quality of our scheme is inferior to that of the approaches based on the subsequent optimization networks (DCAN, Informed, and Fine-tune).

In conclusion, we propose a physics-guided compact single-layer neural network with Φ regularization to train efficient modulation patterns for DGI at a low sampling ratio $\beta = 6.25\%$. Benefiting from the DGI physical process prior, the learned grayscale modulation patterns can be directly used for DGI without any further optimization networks or complex iterative optimization algorithms. We have demonstrated that our scheme is superior in imaging quality compared to some typical GI modulation patterns through simulations and experiments. In addition, we achieve real-time imaging with a video rate of ~ 10 Hz for images with a pixel number of $N = 128 \times 128$. Our scheme complements the existing modulation pattern options and may enhance the performance of computational imaging schemes based on sequential correlation measurements without the use of complex iterative optimization algorithms.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (12274338, 91736104), the Ministry of Science and Technology of the People's Republic of China (2016YFA0301404), and Fundamental Research Funds for the Central Universities. The authors thank Fei Wang (Shanghai Institute of Optics and Fine Mechanics) for meaningful discussions and provision of source code for related work.

The authors declare no conflict of interests.

APPENDIX A: MODULATION PATTERNS PREPROCESSING AND PRINCIPLE OF IMAGE RECONSTRUCTION

In our experiments, we preprocess the modulation patterns to fit the actual imaging system as shown in Fig. 7.

Specifically, in step 1, we divide the m th modulation pattern $H_m(x, y)$ into positive pattern $H_m(x, y)_+$ and negative pattern $H_m(x, y)_-$ according to the values

$$\begin{aligned} H_m(x, y)_+ &= \begin{cases} H_m(x, y), & H_m(x, y) \geq 0, \\ 0, & H_m(x, y) < 0, \end{cases} \\ H_m(x, y)_- &= \begin{cases} -H_m(x, y), & H_m(x, y) \leq 0, \\ 0, & H_m(x, y) > 0, \end{cases} \end{aligned} \quad (\text{A1})$$

where $m = 1, 2, \dots, M$. As stated in the main text, the sampling ratio $\beta = M/N = 6.25\%$, where $M = 1024$ and $N = 16384$.

After considering the positive and negative separation, a total of $2 \times M = 2048$ modulation patterns need to be loaded on the DMD. However, it is well known that DMD has only two states, "ON" and "OFF", to represent "1" and "0", respectively. So our grayscale modulation patterns cannot be directly loaded into DMD yet. We use the Floyd-Steinberg dithering algorithm [74] to dither the above positive and negative patterns (step 3). Before dithering, in step 2, we first calculate the Kronecker product (768×768) of the modulation pattern (128×128) and the all-1s matrix (6×6), and then we obtain the final pattern loaded into the DMD (1024×768) by zero-padding (step 4).

We experimentally compared several typical modulation patterns: random binary patterns (Random), reordered Hadamard patterns (Hadamard), Fourier sinusoidal patterns (Fourier), trained binary patterns (DCAN, Informed, and Fine-tune), and the four sets of learned grayscale modulation patterns obtained by our proposed scheme (Face, Face-Reg, STL10, and STL10-Reg), all with the same sampling ratio of $\beta = 6.25\%$. For Random, Hadamard, and DCAN, since the modulation patterns only take two values of $+1$ and -1 , we did not perform the Floyd-Steinberg dithering (step 3). As for the Fourier modulation patterns, we did not implement the positive and negative pattern separation (step 1), and we use the four-step phase shift method to reconstruct the Fourier spectrum of the target image, which yields a total of $512 \times 4 = 2048$ modulation patterns when the sampling ratio $\beta = 6.25\%$. The modulation patterns of Informed and Fine-tune take two values of $+1$ and 0 , so we did not perform steps 1 and 3. All four sets of our grayscale modulation patterns (Face, Face-Reg, STL10, and STL10-Reg) are preprocessed in the above four steps.

For modulation patterns Random, Face, Face-Reg, STL10, and STL10-Reg, we adopt DGI [11] as the reconstruction algorithm:

$$O_{\text{DGI}} = \langle H_m I_m \rangle - \frac{\langle I_m \rangle}{\langle S_m \rangle} \langle S_m H_m \rangle. \quad (\text{A2})$$

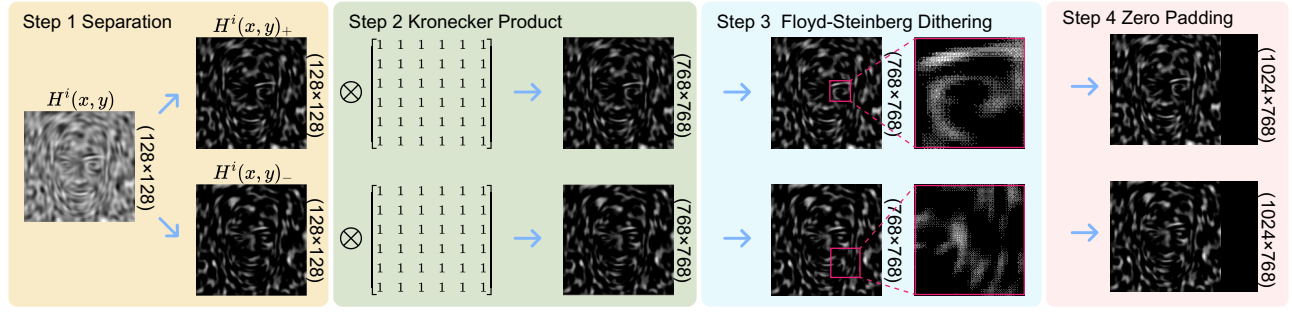


FIG. 7. Four preprocessing steps for modulating patterns in experiments.

Here H_m is the m th modulation pattern, I_m is the m th measurement, S_m is the total intensity of the m th modulation pattern, which is defined as $S_m = \sum_{x,y} H_m(x,y)$, and $\langle \cdot \rangle$ is the ensemble average in terms of m defined as $\langle H_m \rangle = (1/M) \sum_{m=1}^M H_m$ and $\langle I_m \rangle = (1/M) \sum_{m=1}^M I_m$. For the two special sets of orthogonal bases, Hadamard and Fourier, we use the most general reconstruction method as in [75,76]:

$$O_{\text{Hadamard}} = \frac{1}{M} \sum_{m=1}^M H_m I_m = \langle H_m I_m \rangle. \quad (\text{A3})$$

As for the Fourier modulation patterns, we adopt the four-step phase shift method to reconstruct the Fourier spectrum of the target image and reconstruct the object image by applying an inverse Fourier transform [76]. It should be pointed out that we adopt the ‘‘Circular’’ sampling strategy for the Fourier spectrum. For DCAN, Informed, and Fine-tune, we adopt the same subsequent optimization network and iterative optimization Fine-tune process as described in Refs. [63,66].

In our practical experiments, we control the DMD and acquire bucket detection signals using LABVIEW, while image reconstruction is performed using MATLAB due to its superior matrix manipulation speed. Coordinating these two software platforms is complex and time-consuming. Therefore, we adopt a strategy where we first acquire the bucket detection signals (considering the reconstruction time using the DGI algorithm) and then reconstruct the images.

The time required to load 2048 modulation patterns into the DMD memory is on the order of minutes. However, once loaded, the DMD can project these patterns at a display rate of 22 kHz (its maximum operating frequency), resulting in a display time of 0.093 s for the 2048 modulation patterns, which corresponds to a frame rate of 10.74 Hz. Another critical factor is the reconstruction time using DGI, which takes approximately 0.015 s per reconstruction calculation. Considering these two time constraints, we estimate the imaging frame rate of the system to be approximately 9.26 Hz.

Optimizing programming languages and enhancing hardware capabilities could potentially reduce the time required by the reconstruction algorithm. Other time considerations are believed to be negligible or can be addressed through measures such as improving system integration and preloading memory.

APPENDIX B: LEARNED MODULATION PATTERNS

Figure 8 shows some typical samples of modulation patterns trained by the proposed scheme. Figure 8 rows 1, 2, 3, and 4 are selected modulation patterns of Face, Face-Reg, STL10, and STL10-Reg, respectively. It can be seen that modulation patterns trained by the two datasets have distinct characteristics. Face and Face-Reg, trained by the face dataset CelebAMask-HQ, have facial-like features composed of speckles, while STL10 and STL10-Reg, trained on the natural image dataset STL-10, are more like random speckle. Notably, it is these facial features that contribute to the extraordinary performance of face image reconstruction and are responsible for face-like artifacts in nonface reconstruction. We compromise the generalization ability of the modulation patterns to achieve excellent performance in specific scenes, such as face images, which has notable implications.

It is not difficult to see from Fig. 8 that the speckle particle sizes in each set of modulation patterns are well distributed. The modulation pattern with larger speckle particles contributes more to the low-frequency signal of the reconstructed image, and the smaller speckle particles contribute the high-frequency details. Therefore, the proposed scheme can reconstruct the low-frequency signal of the image while retaining more high-frequency details at a very low sampling ratio. Next, we discuss the effect of Φ regularization on modulation patterns. Taking Fig. 8 rows 1 and 3 as controls, it can be concluded by observing Fig. 8 rows 2 and 4 that the speckle particles in the modulation patterns with Φ regularization are more delicate and smooth, which helps to obtain a reconstruction with less noise and artifacts.

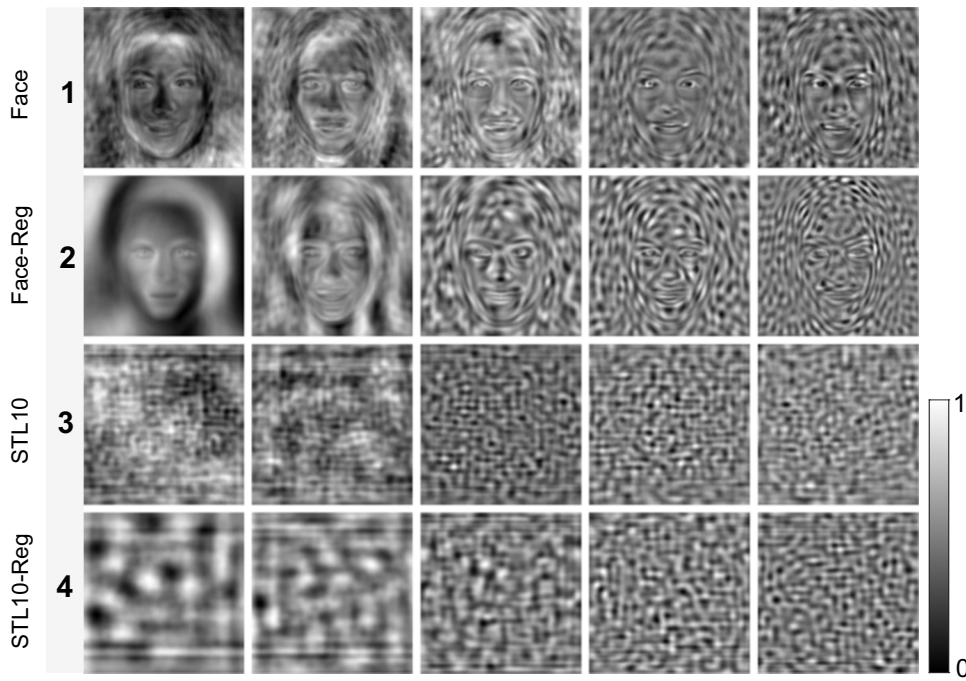


FIG. 8. Normalized typical examples of the learned modulation patterns. Rows 1–4 show some sample modulation patterns for Face, Face-Reg, STL10, and STL10-Reg, respectively. In each row, from left to right, the speckle particles change from large to small.

APPENDIX C: INFLUENCE OF WEIGHT FACTOR λ ON IMAGING RESULTS

We use the weight factor λ to balance the MSE and Φ regularization in the objective function (loss function):

$$\mathbf{H}^* = \operatorname{argmin}_{\mathbf{H}} \|\operatorname{DGI}(\mathbf{H}, \mathbf{HO}) - \mathbf{O}\|^2 + \lambda(D_{\text{var}} + D^{\text{non}})$$

A larger value of λ corresponds to a relatively stronger Φ regularization, and vice versa. We give the Φ matrix with different λ in Fig. 9(b). When $\lambda = 0$, most values of the Φ matrix are nonzero. With the increase of λ , the nondiagonal elements of the Φ matrix gradually converge to zero, and the diagonal elements are more flat. In our simulations and experiments, we empirically chose $\lambda = 0.01$. We test the model with different λ on 20 images consisting of face and natural images.

Figure 9(a) shows the simulated reconstruction results of two sample images (Face4 and Boat) in multiple scenarios. It can be seen that the reconstruction of both Face and STL10 have more distortions and artifacts when $\lambda = 0$, i.e., without Φ regularization. These artifacts decrease and the reconstructed image becomes smoother as λ increases. However, the reconstruction quality of the image does not always grow with λ , which can be seen from the curves of average PSNR and SSIM of the 20 images in Fig. 9(c). The reconstruction quality first increases and then decreases with the increase of λ . This is consistent with the results in Fig. 9(a): as λ increases, the artifacts in the reconstructed

images become fewer, but at the same time the images become more blurred with some details lost.

As mentioned in the main text, there is a theoretical limit for the matrix Φ to approach the identity matrix with a sampling ratio $\beta = 6.25\%$. With Φ regularization, we effectively mitigate the artifacts and noise in the reconstructed images, but also reduce their resolution (i.e., the reconstructed images are more blurred). This is due to the combined effect of the above theoretical limit and the complex competition between MSE and Φ regularization in the objective (loss) function. We illustrate the issue of resolution degradation with the following discussion.

We take the center of the matrix Φ as the origin and establish the coordinate system in the diagonal and subdiagonal directions as shown in Fig. 10(a). In fact, the width of the intensity distribution in the subdiagonal direction (x axis) of the matrix Φ is inversely proportional to the imaging resolution. We crop a total of 1001 line segments with a length of 2049 pixels in the subdiagonal direction, as shown in the red rectangle in Fig. 10(a). Figure 10(b) shows the result of summing these 1001 line segments under different λ cases. For the curve corresponding to each λ in Fig. 10(b), we extract its upper envelope and perform a Gaussian fit to plot its FWHM in Fig. 10(c). Figure 10(c) shows that, with the increase of λ , the FWHM of the upper envelope line increases, which leads to the resolution decrease of the reconstructed image. It can be seen that the addition of Φ regularization reduces the artifacts and noise of the reconstructed image, but also sacrifices a

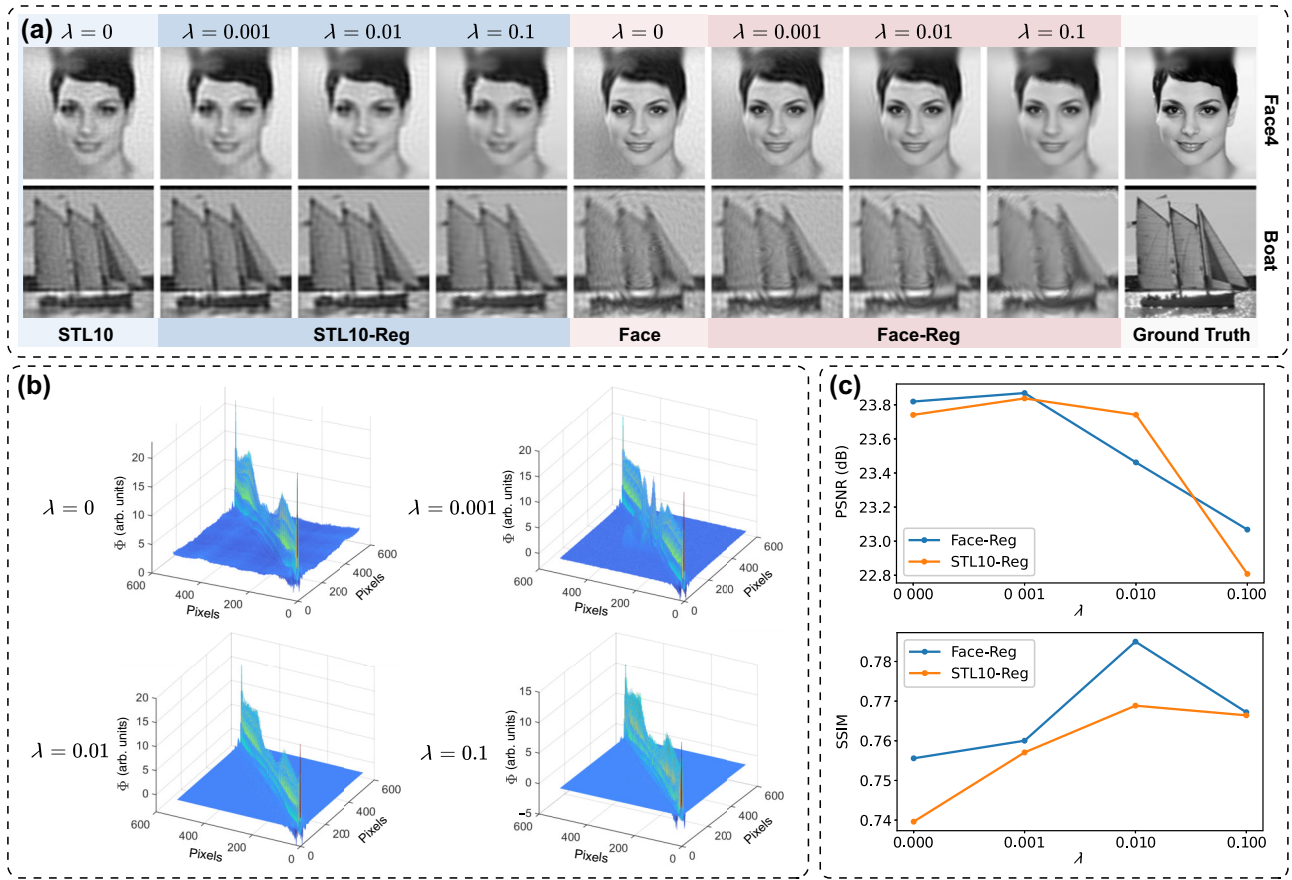


FIG. 9. Effects of different λ on imaging results of the proposed scheme. (a) The reconstruction results of the modulation patterns trained by two datasets with different λ for the imaging targets Face4 and Boat. (b) Three-dimensional display of matrix Φ with $\lambda = 0, 0.001, 0.01,$ and 0.1 . Here, for the purpose of display convenience, the matrix Φ has been downsampled to 512×512 . (c) The curves of average PSNR and SSIM versus λ for the reconstructed results of 20 selected face and natural target images with Face-Reg and STL10-Reg, respectively.

certain resolution, which is consistent with our simulation and experimental results. One has to balance the impact of Φ regularization and make a compromise between fewer

artifacts and higher image resolution. Here, we chose $\lambda = 0.01$ in our simulations and experiments. A better choice for λ may exist, which is beyond the purpose of this work.

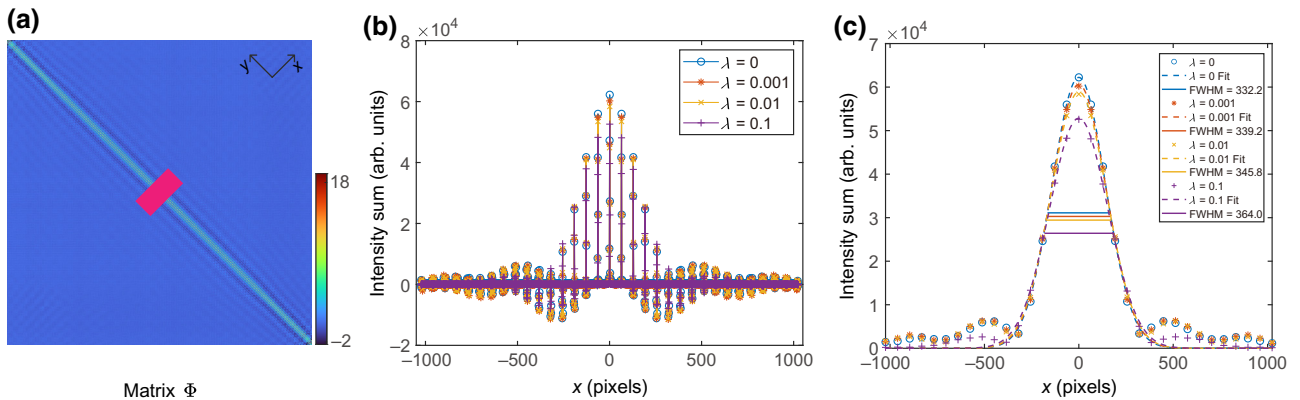


FIG. 10. Numerical simulation analysis of the influence of weight factor λ on imaging resolution. (a) Two-dimensional intensity map of matrix Φ , where the red rectangle is the cropped 1001 line segments with a length of 2049 pixels in the subdiagonal direction. The matrix Φ shown here has been downsampled to facilitate presentation. (b) The curve of summing the 1001 line segments under different λ cases. (c) The extracted upper envelopes of panel (b) and their Gaussian fits with FWHMs indicated by the solid lines.

- [1] T. B. Pittman, Y. Shih, D. Strekalov, and A. V. Sergienko, Optical imaging by means of two-photon quantum entanglement, *Phys. Rev. A* **52**, R3429 (1995).
- [2] R. S. Bennink, S. J. Bentley, and R. W. Boyd, “Two-Photon” Coincidence Imaging with a Classical Source, *Phys. Rev. Lett.* **89**, 113601 (2002).
- [3] A. Gatti, E. Brambilla, M. Bache, and L. A. Lugiato, Correlated imaging, quantum and classical, *Phys. Rev. A* **70**, 013802 (2004).
- [4] A. Valencia, G. Scarcelli, M. D’Angelo, and Y. Shih, Two-Photon Imaging with Thermal Light, *Phys. Rev. Lett.* **94**, 063601 (2005).
- [5] Y.-H. Zhai, X.-H. Chen, D. Zhang, and L.-A. Wu, Two-photon interference with true thermal light, *Phys. Rev. A* **72**, 043805 (2005).
- [6] G. Scarcelli, V. Berardi, and Y. Shih, Phase-conjugate mirror via two-photon thermal light imaging, *Appl. Phys. Lett.* **88**, 061106 (2006).
- [7] J. H. Shapiro, Computational ghost imaging, *Phys. Rev. A* **78**, 061802 (2008).
- [8] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, Single-pixel imaging via compressive sampling, *IEEE Signal Process. Mag.* **25**, 83 (2008).
- [9] R. Meyers, K. S. Deacon, and Y. Shih, Ghost-imaging experiment by measuring reflected photons, *Phys. Rev. A* **77**, 041801 (2008).
- [10] Y. Bromberg, O. Katz, and Y. Silberberg, Ghost imaging with a single detector, *Phys. Rev. A* **79**, 053840 (2009).
- [11] F. Ferri, D. Magatti, L. Lugiato, and A. Gatti, Differential Ghost Imaging, *Phys. Rev. Lett.* **104**, 253603 (2010).
- [12] N. Radwell, K. J. Mitchell, G. M. Gibson, M. P. Edgar, R. Bowman, and M. J. Padgett, Single-pixel infrared and visible microscope, *Optica* **1**, 285 (2014).
- [13] M. P. Edgar, G. M. Gibson, and M. J. Padgett, Principles and prospects for single-pixel imaging, *Nat. Photonics* **13**, 13 (2019).
- [14] Z. Yang, J. Liu, W.-X. Zhang, D. Ruan, and J.-L. Li, Instant single-pixel imaging: On-chip real-time implementation based on the instant ghost imaging algorithm, *OSA Contin.* **3**, 629 (2020).
- [15] I. Hoshi, T. Shimobaba, T. Kakue, and T. Ito, Real-time single-pixel imaging using a system on a chip field-programmable gate array, *Sci. Rep.* **12**, 14097 (2022).
- [16] P. Sen, B. Chen, G. Garg, S. R. Marschner, M. Horowitz, M. Levoy, and H. P. A. Lensch, in *ACM SIGGRAPH 2005 Papers* (Association for Computing Machinery, New York, NY, USA, 2005), p. 745.
- [17] W. L. Chan, K. Charan, D. Takhar, K. F. Kelly, R. G. Baraniuk, and D. M. Mittleman, A single-pixel terahertz imaging system based on compressed sensing, *Appl. Phys. Lett.* **93**, 121105 (2008).
- [18] L. Bian, J. Suo, G. Situ, Z. Li, J. Fan, F. Chen, and Q. Dai, Multispectral imaging using a single bucket detector, *Sci. Rep.* **6**, 1 (2016).
- [19] Z. Zhang, S. Liu, J. Peng, M. Yao, G. Zheng, and J. Zhong, Simultaneous spatial, spectral, and 3D compressive imaging via efficient Fourier single-pixel measurements, *Optica* **5**, 315 (2018).
- [20] F. Magalhães, F. M. Araújo, M. Correia, M. Abolbashari, and F. Farahi, High-resolution hyperspectral single-pixel imaging system based on compressive sensing, *Opt. Eng.* **51**, 071406 (2012).
- [21] C. M. Watts, D. Shrekenhamer, J. Montoya, G. Lipworth, J. Hunt, T. Sleasman, S. Krishna, D. R. Smith, and W. J. Padilla, Terahertz compressive imaging with metamaterial spatial light modulators, *Nat. Photonics* **8**, 605 (2014).
- [22] R. I. Stantchev, B. Sun, S. M. Hornett, P. A. Hobson, G. M. Gibson, M. J. Padgett, and E. Hendry, Noninvasive, near-field terahertz imaging of hidden objects using a single-pixel detector, *Sci. Adv.* **2**, e1600190 (2016).
- [23] L. Olivieri, J. S. T. Gongora, L. Peters, V. Cecconi, A. Cutrona, J. Tunesi, R. Tucker, A. Pasquazi, and M. Peccianti, Hyperspectral terahertz microscopy via nonlinear ghost imaging, *Optica* **7**, 186 (2020).
- [24] J. S. Toterogongora, L. Olivieri, L. Peters, J. Tunesi, V. Cecconi, A. Cutrona, R. Tucker, V. Kumar, A. Pasquazi, and M. Peccianti, Route to intelligent imaging reconstruction via terahertz nonlinear ghost imaging, *Micromachines* **11**, 521 (2020).
- [25] R. I. Stantchev, X. Yu, T. Blu, and E. Pickwell-MacPherson, Real-time terahertz imaging with a single-pixel detector, *Nat. Commun.* **11**, 2535 (2020).
- [26] H. Yu, R. Lu, S. Han, H. Xie, G. Du, T. Xiao, and D. Zhu, Fourier-Transform Ghost Imaging with Hard X Rays, *Phys. Rev. Lett.* **117**, 113901 (2016).
- [27] D. Pelliccia, A. Rack, M. Scheel, V. Cantelli, and D. M. Paganin, Experimental X-Ray Ghost Imaging, *Phys. Rev. Lett.* **117**, 113902 (2016).
- [28] A.-X. Zhang, Y.-H. He, L.-A. Wu, L.-M. Chen, and B.-B. Wang, Tabletop X-ray ghost imaging with ultra-low radiation, *Optica* **5**, 374 (2018).
- [29] C. Wang, X. Mei, L. Pan, P. Wang, W. Li, X. Gao, Z. Bo, M. Chen, W. Gong, and S. Han, Airborne near infrared three-dimensional ghost imaging lidar via sparsity constraint, *Remote Sens. (Basel)* **10**, 732 (2018).
- [30] C. Zhao, W. Gong, M. Chen, E. Li, H. Wang, W. Xu, and S. Han, Ghost imaging lidar via sparsity constraints, *Appl. Phys. Lett.* **101**, 141123 (2012).
- [31] O. S. Magana-Loaiza, G. A. Howland, M. Malik, J. C. Howell, and R. W. Boyd, Compressive object tracking using entangled photons, *Appl. Phys. Lett.* **102**, 231104 (2013).
- [32] E. Li, Z. Bo, M. Chen, W. Gong, and S. Han, Ghost imaging of a moving target with an unknown constant speed, *Appl. Phys. Lett.* **104**, 251120 (2014).
- [33] J. Wu, L. Hu, and J. Wang, Fast tracking and imaging of a moving object with single-pixel imaging, *Opt. Express* **29**, 42589 (2021).
- [34] B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. J. Padgett, 3D computational imaging with single-pixel detectors, *Science* **340**, 844 (2013).
- [35] M.-J. Sun, M. P. Edgar, G. M. Gibson, B. Sun, N. Radwell, R. Lamb, and M. J. Padgett, Single-pixel three-dimensional imaging with time-based depth resolution, *Nat. Commun.* **7**, 1 (2016).
- [36] K. Ota and Y. Hayasaki, Complex-amplitude single-pixel imaging, *Opt. Lett.* **43**, 3682 (2018).

- [37] R. Liu, S. Zhao, P. Zhang, H. Gao, and F. Li, Complex wavefront reconstruction with single-pixel detector, *Appl. Phys. Lett.* **114**, 161901 (2019).
- [38] S. Zhao, R. Liu, P. Zhang, H. Gao, and F. Li, Fourier single-pixel reconstruction of a complex amplitude optical field, *Opt. Lett.* **44**, 3278 (2019).
- [39] S. Zhao, S. Chen, X. Wang, R. Liu, P. Zhang, H. Li, H. Gao, and F. Li, Measuring the complex spectrum of orbital angular momentum and radial index with a single-pixel detector, *Opt. Lett.* **45**, 5990 (2020).
- [40] E. Tajahuerce, V. Durán, P. Clemente, E. Irlés, F. Soldevila, P. Andrés, and J. Lancis, Image transmission through dynamic scattering media by single-pixel photodetection, *Opt. Express* **22**, 16945 (2014).
- [41] V. Durán, F. Soldevila, E. Irlés, P. Clemente, E. Tajahuerce, P. Andrés, and J. Lancis, Compressive imaging in scattering media, *Opt. Express* **23**, 14424 (2015).
- [42] K. L. C. Seow, P. Török, and M. R. Foreman, Single pixel polarimetric imaging through scattering media, *Opt. Lett.* **45**, 5740 (2020).
- [43] N. Huynh, E. Zhang, M. Betcke, S. Arridge, P. Beard, and B. Cox, Single-pixel optical camera for video rate ultrasonic imaging, *Optica* **3**, 26 (2016).
- [44] J. Yang, L. Gong, X. Xu, P. Hai, Y. Shen, Y. Suzuki, and L. V. Wang, Motionless volumetric photoacoustic microscopy with spatially invariant resolution, *Nat. Commun.* **8**, 1 (2017).
- [45] Y. Guo, B. Li, and X. Yin, Single-shot compressed photoacoustic tomographic imaging with a single detector in a scattering medium, *Phys. Rev. Appl.* **13**, 044009 (2020).
- [46] X. Zhang, S. Song, X. Ma, H. Zhang, L. Gai, Y. Gu, and W. Li, Optimizing ghost imaging via analysis and design of speckle patterns, *Appl. Opt.* **61**, 4113 (2022).
- [47] M.-J. Sun, L.-T. Meng, M. P. Edgar, M. J. Padgett, and N. Radwell, A Russian dolls ordering of the Hadamard basis for compressive single-pixel imaging, *Sci. Rep.* **7**, 1 (2017).
- [48] W.-K. Yu and Y.-M. Liu, Single-pixel imaging with origami pattern construction, *Sensors* **19**, 5135 (2019).
- [49] W.-K. Yu, Super sub-Nyquist single-pixel imaging by means of cake-cutting Hadamard basis sort, *Sensors* **19**, 4122 (2019).
- [50] L. López-García, W. Cruz-Santos, A. García-Arellano, P. Filio-Aguilar, J. A. Cisneros-Martínez, and R. Ramos-García, Efficient ordering of the Hadamard basis for single pixel imaging, *Opt. Express* **30**, 13714 (2022).
- [51] Z. Zhang, X. Ma, and J. Zhong, Single-pixel imaging by means of fourier spectrum acquisition, *Nat. Commun.* **6**, 1 (2015).
- [52] K. Guo, S. Jiang, and G. Zheng, Multilayer fluorescence imaging on a single-pixel detector, *Biomed. Opt. Express* **7**, 2425 (2016).
- [53] O. Katz, Y. Bromberg, and Y. Silberberg, Compressive ghost imaging, *Appl. Phys. Lett.* **95**, 131110 (2009).
- [54] W.-K. Yu, M.-F. Li, X.-R. Yao, X.-F. Liu, L.-A. Wu, and G.-J. Zhai, Adaptive compressive ghost imaging based on wavelet trees and sparse representation, *Opt. Express* **22**, 7133 (2014).
- [55] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).
- [56] M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, Deep-learning-based ghost imaging, *Sci. Rep.* **7**, 1 (2017).
- [57] Y. He, G. Wang, G. Dong, S. Zhu, H. Chen, A. Zhang, and Z. Xu, Ghost imaging based on deep learning, *Sci. Rep.* **8**, 1 (2018).
- [58] S. Rizvi, J. Cao, K. Zhang, and Q. Hao, Improving imaging quality of real-time Fourier single-pixel imaging via deep learning, *Sensors* **19**, 4190 (2019).
- [59] F. Wang, H. Wang, H. Wang, G. Li, and G. Situ, Learning from simulation: An end-to-end deep-learning approach for computational ghost imaging, *Opt. Express* **27**, 25560 (2019).
- [60] I. Hoshi, T. Shimobaba, T. Kakue, and T. Ito, Single-pixel imaging using a recurrent neural network combined with convolutional layers, *Opt. Express* **28**, 34069 (2020).
- [61] Z. Zhang, X. Li, S. Zheng, M. Yao, G. Zheng, and J. Zhong, Image-free classification of fast-moving objects using “learned” structured illumination and single-pixel detection, *Opt. Express* **28**, 13269 (2020).
- [62] H. Fu, L. Bian, and J. Zhang, Single-pixel sensing with optimal binarized modulation, *Opt. Lett.* **45**, 3111 (2020).
- [63] C. F. Higham, R. Murray-Smith, M. J. Padgett, and M. P. Edgar, Deep learning for real-time single-pixel video, *Sci. Rep.* **8**, 1 (2018).
- [64] I. Hoshi, T. Shimobaba, T. Kakue, and T. Ito, Optimized binary patterns by gradient descent for ghost imaging, *IEEE Access* **9**, 97320 (2021).
- [65] X. Zhan, L. Bian, C. Zhu, and J. Zhang, Weighted encoding optimization for dynamic single-pixel imaging and sensing, *arXiv:2201.02833*.
- [66] F. Wang, C. Wang, C. Deng, S. Han, and G. Situ, Single-pixel imaging using physics enhanced deep learning, *Photonics Res.* **10**, 104 (2022).
- [67] R. A. Aguilar and D. Dailisan, Deep-learned orthogonal basis patterns for fast, noise-robust single-pixel imaging, *arXiv:2205.08736*.
- [68] C.-H. Lee, Z. Liu, L. Wu, and P. Luo, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), p. 5549.
- [69] A. Coates, A. Ng, and H. Lee, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, edited by G. Gordon, D. Dunson and M. Dudik (PMLR, Fort Lauderdale, FL, USA, 2011), Vol. 15, p. 215, <https://proceedings.mlr.press/v15/coates11a.html>.
- [70] <https://github.com/emrysw/DGI-SLNN>.
- [71] J. W. Gibbs, Fourier’s series, *Nature* **59**, 200 (1898).
- [72] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, and D. Psaltis, Inference in artificial intelligence with deep optics and photonics, *Nature* **588**, 39 (2020).
- [73] See Supplemental Material at <https://link.aps.org/supplemental/10.1103/PhysRevApplied.22.014023> for visualization

- 1: Real-time imaging results of the test scene using the SLT10-Reg model. The scene consists of a dynamic toy astronaut and part of the USAF resolution chart as the background.
- [74] P. D. Wellner, Adaptive thresholding for the DigitalDesk, Xerox, EPC1993-110 **404** (1993).
- [75] G. M. Gibson, S. D. Johnson, and M. J. Padgett, Single-pixel imaging 12 years on: A review, *Opt. Express* **28**, 28190 (2020).
- [76] Z. Zhang, X. Wang, G. Zheng, and J. Zhong, Hadamard single-pixel imaging versus Fourier single-pixel imaging, *Opt. Express* **25**, 19619 (2017).