PHOTONICS Research

Deep compressed imaging via optimized pattern scanning

KANGNING ZHANG, JUNJIE HU, AND WEIJIAN YANG*

Department of Electrical and Computer Engineering, University of California, Davis, California 95616, USA
**Corresponding author: wejyang@ucdavis.edu*

Received 7 October 2020; revised 31 December 2020; accepted 13 January 2021; posted 15 January 2021 (Doc. ID 410556); published 1 March 2021

The need for high-speed imaging in applications such as biomedicine, surveillance, and consumer electronics has called for new developments of imaging systems. While the industrial effort continuously pushes the advance of silicon focal plane array image sensors, imaging through a single-pixel detector has gained significant interest thanks to the development of computational algorithms. Here, we present a new imaging modality, deep compressed imaging via optimized-pattern scanning, which can significantly increase the acquisition speed for a single-detector-based imaging system. We project and scan an illumination pattern across the object and collect the sampling signal with a single-pixel detector. We develop an innovative end-to-end optimized auto-encoder, using a deep neural network and compressed sensing algorithm, to optimize the illumination pattern, which allows us to reconstruct faithfully the image from a small number of measurements, with a high frame rate. Compared with the conventional switching-mask-based single-pixel camera and point-scanning imaging systems, our method achieves a much higher imaging speed, while retaining a similar imaging quality. We experimentally validated this imaging modality in the settings of both continuous-wave illumination and pulsed light illumination and showed high-quality image reconstructions with a high compressed sampling rate. This new compressed sensing modality could be widely applied in different imaging systems, enabling new applications that require high imaging speeds.

https://doi.org/10.1364/PRJ.410556

1. INTRODUCTION

High-speed imaging has become more and more crucial in many new applications, such as in biomedicine, surveillance, and consumer electronics. There are two roadmaps for high-speed optical imaging: engineering a faster focal plane array image sensor and developing new imaging modalities using a single-pixel detector. Although intense industrial efforts have been made, high-speed and low-noise silicon focal plane array cameras are still expensive. Furthermore, imaging at wavelengths outside the silicon sensitivity spectrum can make the focal plane array cameras considerably more complicated [1]. In contrast, imaging through a single-pixel detector, which shrinks a photodetector array down to a single unit [2–9], can enormously reduce the cost and offer additional features such as reduced pixel crosstalk. A popular imaging modality of the single-pixel detector is based on point scanning [5–9] [Fig. 1(a)], for example, the laser-scanning microscopes that are commonly used in biomedicine [5–8]. However, such methods are speed limited due to the point-by-point data acquisition. Another approach with the single-pixel detector relies on compressed sensing (CS) [10–12], represented by the switching-mask-based single-pixel cameras [2–4] [Fig. 1(b)]. There, the images are encoded by a series of spatially well-designed sampling patterns. For each pattern, all pixels across the entire image are summed and collected by the detector. Leveraging the general prior knowledge of sparsity in images, CS is used to reconstruct the image through a small number of measurements [2–4,13,14]. Although the sampling rate can be below the Nyquist criterion, the imaging frame rate is limited by how fast the sampling pattern can be switched and cycled, which is typically conducted by a digital micromirror device (DMD) [15] and operates at <22.7 kHz. Using an LED array to generate the pattern could increase the overall speed, but so far only 32 × 32 pixel images have been demonstrated [16], and it may be expensive to scale up. It does not allow passive light illumination (i.e., structured detection) either.

In this paper, we propose and demonstrate a deep compressed sensing modality, which can significantly increase the imaging speed while preserving a high reconstruction quality. This approach combines the strength of both compressed sensing and point-scanning imaging, and we call it deep compressed imaging via optimized pattern scanning (DeCIOPS). Instead of projecting multiple binary patterns onto the entire object sequentially, we utilize only one gray-scale optimized

pattern and project it to a small subset of the object. We then scan the pattern across the object by using fast scanning mirrors and collect the signal convolutionally using a single-pixel detector [Fig. 1(c)]. Compared with the conventional single-pixel camera, which relies on sequentially switching the sampling pattern on a DMD, our scanning approach significantly increases the sampling speed. Compared with the point-scanning system, our method samples a much larger portion of the object at once and recovers the resolution computationally. This allows a great reduction of the sampling number and thus increases the frame rate. We note that the improvement of imaging speed does not require an increase of light energy. In fact, the required light dosage in our method is smaller than in the conventional point-scanning system due to a reduced number of measurements. We build an auto-encoder framework [17] to optimize the sampling pattern. The image acquisition system is treated as an encoder, where the high-resolution object is encoded through the sampling pattern into a few measurements. We then formulate an iterative shrinkage-thresholding algorithm network (ISTA-Net) [18], a CS-induced neural network inspired by the iterative shrinkage-thresholding algorithm (ISTA) [19], as a decoder to reconstruct the image. This auto-encoder is trained in an end-to-end fashion. Such a framework can learn an optimized sampling pattern and simultaneously recover a high-resolution image by extracting the feature of sparsity and searching the optimal pair of the encoder and decoder with the lowest incoherence [11], which is one key feature of DeCIOPS versus other compressed sensing or deep-learning-based super-resolution imaging modalities [20–23]. This new imaging modality can be accustomed to any light-scanning imaging system and will greatly benefit the high-throughput imaging applications.

This paper is organized as follows. In Section 2, we introduce the mechanisms of the imaging modality in two configurations of illuminations, continuous-wave (CW) and pulsed light source, as well as the auto-encoder framework and the deep compressed sensing neural network for optimizing the imaging and reconstruction. In Sections 3 and 4, we show the simulation results and experimental results. In Section 5, we discuss the system performance under different signal-to-noise ratio (SNR) and compression ratio and how DeCIOPS can be applied in two-photon microscopy and passive lighting conditions.

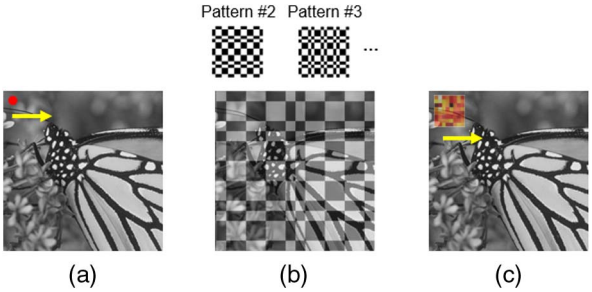

Fig. 1. Landscape of imaging methods using a single-pixel detector. (a) Point scanning system where the signal from an individual pixel is sequentially recorded. (b) A conventional single-pixel camera where different patterns are sequentially projected on the entire object, and the overlap integrals between the object and each pattern are measured. (c) Deep compressed imaging via optimized pattern scanning (DeCIOPS), where a pattern is scanned across the object, and the subsampled convolution between the pattern and the object is measured.

2. PRINCIPLE

A. Image Formation

In DeCIOPS, an illumination pattern is generated through a mask and is scanned across the object by using a set of scanning mirrors. The detector records a subsample of the two-dimensional (2D) convolution between the pattern and the object (Fig. 2). Similar to any point-scanning imaging system, DeCIOPS can use either CW or pulsed light sources. The former is commonly used in imaging systems, whereas the latter is specialized for nonlinear microscopy. When a CW light source or high-repetition-rate pulsed light source is used, the detector continuously integrates the signal as the pattern scans. Hence, in DeCIOPS, we project a rectangle-shaped pattern to the object. By finely adjusting the integration time, each acquisition measures a square subset of the object with a desired resolution [Fig. 2(c)]. In case of a low-repetition-rate pulsed light source, the sampling time stamps of the detector are synchronized with the pulse train, and a square-shaped pattern is projected to the object. By matching the spatial sampling step with the size of the pattern, the entire object is sampled appropriately [Fig. 2(d)].

Mathematically, in low-repetition-rate pulsed light illumination, the image formation of DeCIOPS can be expressed as

$$\boldsymbol{b} = f_{n\times n}(\boldsymbol{x} * \boldsymbol{g}), \tag{1}$$

where $\boldsymbol{x}$ is the object, $\boldsymbol{g}$ is a square-shaped illumination pattern, $*$ represents the 2D convolution, $f_{n\times n}(\cdot)$ models the $n \times n$ undersampling, and $\boldsymbol{b}$ is the measured image. Here, we assume that the mask has a size of $n \times n$ pixels. In the particular case

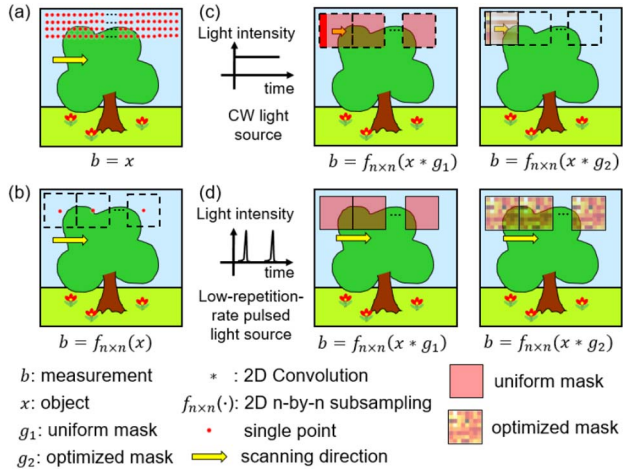

Fig. 2. Schematic of the undersampling schemes in DeCIOPS. (a) Conventional pixel-by-pixel point scanning. (b) Pixel-by-pixel point scanning with a simple undersampling scheme. (c) DeCIOPS in a CW light source configuration with an illumination pattern of a uniform mask (left) or an optimized mask (right). (d) DeCIOPS in a low-repetition-rate pulsed light source configuration with a uniform mask (left) or an optimized mask (right) as an illumination pattern. The mathematic formula below each panel illustrates the process of image formation, where $\boldsymbol{g_1}$ and $\boldsymbol{g_2}$ are both square shaped.

where $\boldsymbol{g}$ is uniform mask $\boldsymbol{g}_1$ [Fig. 2(d), left], Eq. (1) is equivalent to a naïve undersampling by unweighted averaging of every $n \times n$ pixels of the full resolution image [Fig. 2(a)] acquired in single-point scanning. As discussed in Section 2.B, $\boldsymbol{g}$ can be optimized to achieve the best image reconstruction performance [Fig. 2(d), right].

In the CW light or high-repetition-rate pulsed light illumination case, where the detector continuously integrates the signal, we configure the illumination pattern in a size of $n \times 1$ pixels. When this pattern sweeps n columns, the information of $n \times n$ pixels is integrated into a single measurement [Fig. 2(c)]. We can use the same mathematical formulation as Eq. (1) to model the image formation, where each column in the $n \times n$ mask $\boldsymbol{g}$ is identical.

B. End-to-End Optimized Auto-Encoder Framework

In DeCIOPS, we build an auto-encoder framework to simultaneously learn the optimized mask pattern and a neural network for image reconstruction (Fig. 3). The encoder block models the image formation through the following expression:

$$\boldsymbol{b}^+ = f_{n\times n}(\boldsymbol{x} * \boldsymbol{g}) + \boldsymbol{\epsilon} = F(\boldsymbol{x}) + \boldsymbol{\epsilon} = \Phi\boldsymbol{x} + \boldsymbol{\epsilon}, \tag{2}$$

where F is an operator, Φ is the linear transfer matrix, both equivalent to the subsampled convolution with mask $\boldsymbol{g}$ in Eq. (1), $\boldsymbol{\epsilon}$ is the additive noise inherent in the imaging system, and $\boldsymbol{b}^+$ is the measured mask-encoded image.

The decoder takes $\boldsymbol{b}^+$ as the input and aims to reconstruct the original object $\boldsymbol{x}$ by solving the following convex optimization problem with a constraint of the sparse representation of $\boldsymbol{x}$:

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\|\Phi\boldsymbol{x} - \boldsymbol{b}^+\|_2^2 + \lambda\|\Psi(\boldsymbol{x})\|_1, \tag{3}$$

where $\Psi(\boldsymbol{x})$ denotes a transform of $\boldsymbol{x}$ into a sparse representation under the basis of Ψ, and λ is a hyperparameter.

The solution of the problem in Eq. (3) can be initialized by calculating the pseudo-inverse of the encoder from the measurement $\boldsymbol{b}^+$. We then use ISTA [19] to find an optimized solution of $\boldsymbol{x}$ by iterating the following two steps:

$$\boldsymbol{r}^{(k)} = \boldsymbol{x}^{(k-1)} - \rho\Phi^T(\Phi\boldsymbol{x}^{(k-1)} - \boldsymbol{b}^+), \tag{4}$$

$$\boldsymbol{x}^{(k)} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{r}^{(k)}\|_2^2 + \lambda\|\Psi(\boldsymbol{x})\|_1, \tag{5}$$

where k denotes the kth ISTA iteration step, and ρ is the step size.

As $\Psi(\cdot)$ is predefined empirically and may not be suitable for the data, we adopt the ISTA-Net algorithm [18], which can learn $\Psi(\cdot)$ through the data. In ISTA-Net, $\Psi(\cdot)$ is replaced with a trainable neural network $\mathcal{F}(\cdot)$, and the optimization problem in Eq. (3) can be rephrased into the following L1-norm regularization problem with a nonlinear transform $\mathcal{F}$:

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\|\Phi\boldsymbol{x} - \boldsymbol{b}^+\|_2^2 + \lambda\|\mathcal{F}(x)\|_1. \tag{6}$$

The kth iteration step in the original ISTA is replaced by a series of symmetric learnable parameters in the kth ISTA-Net phase:

$$\mathcal{F}(\boldsymbol{x}^{(k)}) = \text{soft}(\mathcal{F}(\boldsymbol{r}^{(k)}), \theta), \tag{7}$$

$$\boldsymbol{x}^{(k)} = \mathcal{F}^{-1}(\text{soft}(\mathcal{F}(\boldsymbol{r}^{(k)}), \theta)), \tag{8}$$

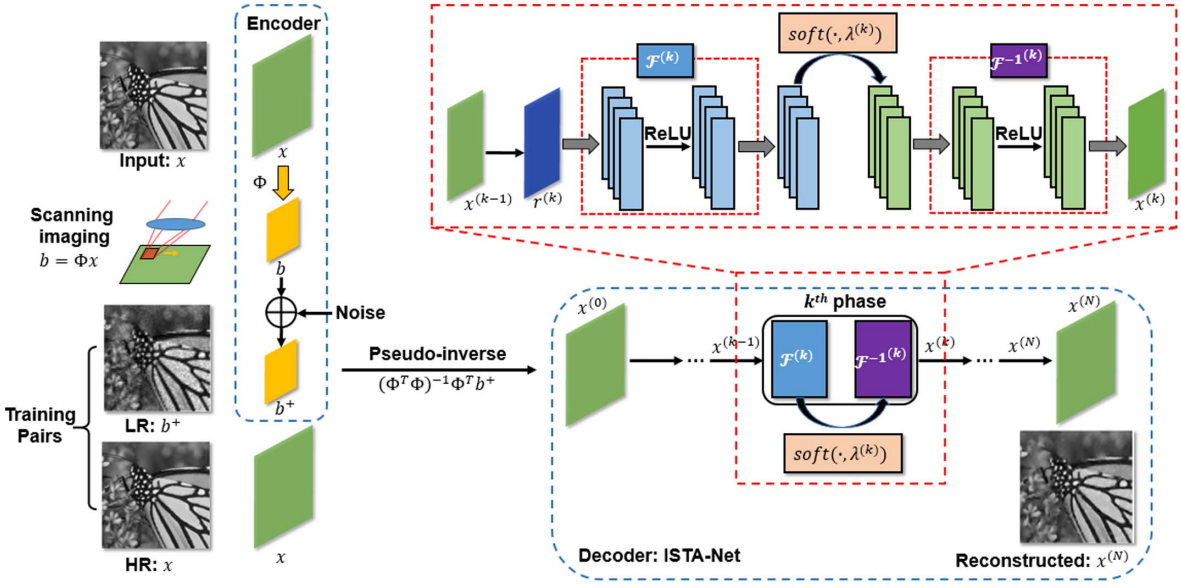

Fig. 3. End-to-end optimized auto-encoder framework of image formation and reconstruction in DeCIOPS. The encoder models the image formation. It encodes the high-resolution (HR) object $\boldsymbol{x}$ into a low-resolution (LR) output $\boldsymbol{b}^+$ through subsampled convolution $\boldsymbol{\Phi}$ and additive noise. The decoder is implemented with an ISTA-Net, which contains N phases and reconstructs the object $\boldsymbol{x}^{(N)}$. Each phase is realized by a structure-symmetric pair of a forward transform $\mathcal{F}^{(k)}$ and a backward transform $\mathcal{F}^{-1(k)}$ with a soft shrinkage threshold, which factually matches one iteration in the conventional ISTA. ReLU, rectified linear unit; Soft(·), soft shrinkage threshold.

where θ is a learnable parameter in the kth module, $\mathcal{F}^{-1}(\cdot)$ is the inverse of $\mathcal{F}(\cdot)$, and soft$(\cdot)$ represents the soft shrinkage threshold. Finally, we obtain the output of the decoder $\boldsymbol{x}^{(N)}$ after a total number of N ISTA-Net phases.

The loss function of ISTA-Net is obtained by calculating the mean square error (MSE) between the output and the ground truth with the constraint of $\mathcal{F}^{-1} \circ \mathcal{F} = I$, as both $\mathcal{F}^{-1}$ and $\mathcal{F}$ are learnable and symmetric, where I is the identity operator. As a result, we have the following loss function with the symmetry constraint:

$$L_{\text{total}} = L_{\text{error}} + \gamma L_{\text{constraint}} = \|\boldsymbol{x}^{(N)} - \boldsymbol{x}\|_2^2 - \gamma\left(\sum_{k=1}^{N} \|\mathcal{F}^{-1(k)}(\mathcal{F}^{(k)}(\boldsymbol{x})) - \boldsymbol{x}\|_2^2\right), \quad (9)$$

where γ is the weight of the symmetry constraint.

3. SIMULATION RESULTS

We trained the auto-encoder using 1500 samples of natural scenes (2D gray-scale image, 256×256 pixel size) from ImageNet [24] and validated the model with 79 samples from two widely used benchmark datasets: Set11 [25] and BSD68 [26]. As an illustration, we chose an undersampling rate of 6.25% (4×4 undersampling) and initialized the pattern $\boldsymbol{g}$ as a 4×4 normalized random Gaussian matrix. γ was set to be 0.01 in the loss function, accompanied by Adam optimization with a learning rate of 1×10^{-4}. We included additive noises in the measurement (5%–10% of the signal strength). The training was performed on a GPU RTX2080Ti 11 GB. The training work of $N = 9$ ISTA-Net phases takes ~5 h for 200 epochs with a batch size of five. We trained two independent auto-encoders, one with a constraint on $\boldsymbol{g}$ so each column of $\boldsymbol{g}$ is identical (CW light or high-repetition-rate pulsed light illumination) and one without such a constraint (low-repetition-rate

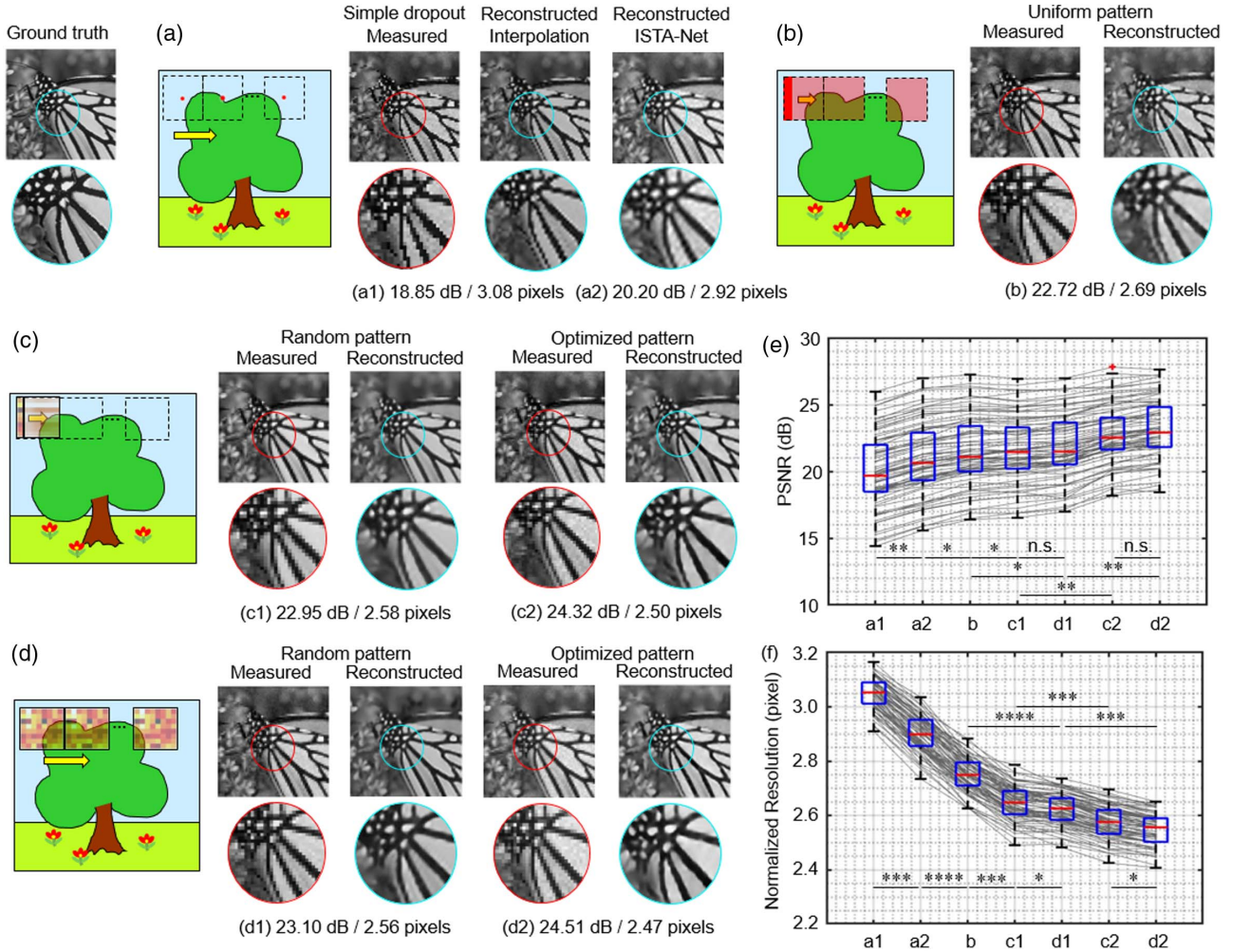

Fig. 4. Comparison of the reconstruction performance in the validation data set Set11 and BSD68, at an undersampling rate of 6.25%, through (a) a simple dropout, (b) an unweighted average (uniform pattern), (c) a random or an optimized illumination pattern (DeCIOPS) with a constraint of identical column, and (d) a random or an optimized illumination pattern (DeCIOPS). The PSNR and resolution of the reconstructed images are labeled below the exemplary sample. (e) PSNR of the reconstructed images of all 79 samples in the validation dataset for cases in (a)–(d). (f) Resolution of the reconstructed images of all 79 samples in the validation dataset for cases in (a)–(d). n.s., not significant; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$, in one-way analysis of variance (ANOVA).

pulsed light illumination). The reconstruction results are evaluated in terms of peak signal-to-noise ratio (PSNR) and spatial resolution by using Fourier ring correlation [27] in the validation dataset. As a control measure, we compared the reconstruction performance of the optimized pattern [Figs. 2(c) and 2(d), right] with a random pattern and two naïve undersampling schemes of the full resolution image either through an unweighted averaging of 4×4 pixels [equivalent to the uniform pattern, Figs. 2(c) and 2(d), left] or a simple dropout [i.e., pick one pixel in every 4×4 and drop out the others, Fig. 2(b)]. All of the simulation groups employ an independently trained ISTA-Net for image reconstruction. In addition, we also used B-spline interpolation [28] to reconstruct the image that was undersampled through a simple dropout.

We compared the measurements and their corresponding reconstruction results for various schemes (Fig. 4). In all cases, the reconstructed results can resolve higher resolution features than the raw measurement. In the undersampling through simple dropout [Fig. 4(a)], ISTA-Net shows better performance than the interpolation. These results demonstrate the effectiveness of ISTA-Net. As the undersampling through dropout [Fig. 4(a)] misses a substantial amount of information in the original object, its reconstruction result is expected to be the worst. Comparing the undersampling through the uniform pattern [Fig. 4(b)] and the optimized pattern [Figs. 4(c) and 4(d)], we find that the optimized pattern cases show reconstruction results with sharper edges. We reason that this is because the optimized pattern balances both the high-frequency and low-frequency components of the original object during sampling, whereas the uniform pattern performs a low-pass filtering such that the high-frequency component is lost before being reconstructed by the decoder. Indeed, when comparing the quantitative results of the PSNR [Fig. 4(e)] and the resolution of the reconstructed image through Fourier ring correlation [Fig. 4(f)], we find that the optimized pattern case shows the best performance. It is important to note that the optimized pattern also outperforms the exemplary random pattern [Figs. 4(c) and 4(d)] for both PSNR and resolution. This verifies the effectiveness of our end-to-end optimized auto-encoder framework. We also note that the optimized mask without the constraints of identical columns show a better performance in resolution than the one with constraints, though their PSNR does not show a significant difference.

4. EXPERIMENTAL RESULTS

To validate the numerical simulation results, we built an imaging system for DeCIOPS (Fig. 5, more details in Appendix A). Here, we used a DMD to generate the light pattern, as it offers great flexibility in comparing the performance between different patterns. In general, as DeCIOPS requires only one illumination pattern, a fixed pattern mask can be used. To generate a gray-scale mask from the DMD binary pixels, we binned 32×32 pixels in the DMD into a super-pixel, which could provide up to 1025 gray-scale levels. A total of 4×4 super-pixels were programmed to generate the optimized mask pattern. A 520 nm diode CW laser source was expanded in beam size and collimated to illuminate the mask pattern on the DMD. The spatially encoded light then passed a $4f$ system composed of a tube lens and an objective lens to reduce its beam size. The light pattern was then raster scanned across the sample through an optically coupled resonant-galvo scanner set. We used a photolithography mask as the sample. The light transmitting through the sample was then collected by a photodetector through a collection lens. While we built this transmission-based imaging system for simplicity, we could turn it into a reflection-based system by adding a beam splitter in front of the sample. We used ScanImage [29] as control software for data acquisition. By adding another $4f$ system composed of cylindrical lenses right after the objective lens, we could turn the 4×4 size pattern into 4×1 size (Appendix A.1). Thus, we could use the same setup to validate the different sampling schemes.

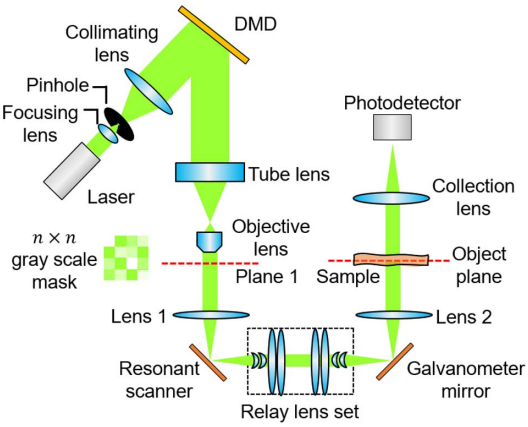

Fig. 5. Experiment setup of DeCIOPS. The laser beam is spatially filtered to improve its spatial uniformity and symmetricity, collimated and expanded in size, and then incident onto a DMD. The beam is spatially modulated by the DMD and then shrunk in size by a $4f$ system formed by a tube lens and an objective lens. The light pattern is scanned by a resonant-galvo scanner set, where a resonant scanner and a galvanometer mirror are optically coupled through a relay lens set. The transmitted light from the sample is collected by a photodetector through a collection lens. The $n \times n$ pattern is generated by the DMD. With an additional $4f$ system with cylindrical lenses after the objective lens, the $n \times n$ pattern can be turned into $n \times 1$ size (Appendix A.1). The red dashed line (plane 1 and object plane) indicates the conjugate plane of the gray-scale pattern mask.

As the optical mode from the diode laser did not have uniform intensity, and the pattern could be corrupted by laser interference, we calibrated the DMD to ensure the illumination pattern on the sample plane matched well with the design (Appendix A.3). While our imaging system is naturally a CW light imaging system, we could also mimic the pulsed light source condition through an additional digital sampling step after the image acquisition (Appendix A.5).

A. Reconstruction Results with a CW Light Source

We evaluated the experimental results in the natural CW light source setting with the 4×1 pattern mask. We compared the reconstruction results across the four undersampling schemes: a simple dropout, the uniform illumination pattern, an exemplary random illumination pattern, and the optimized illumination pattern for various samples (Fig. 6). The ground truth images were obtained by the high-resolution point-scanning

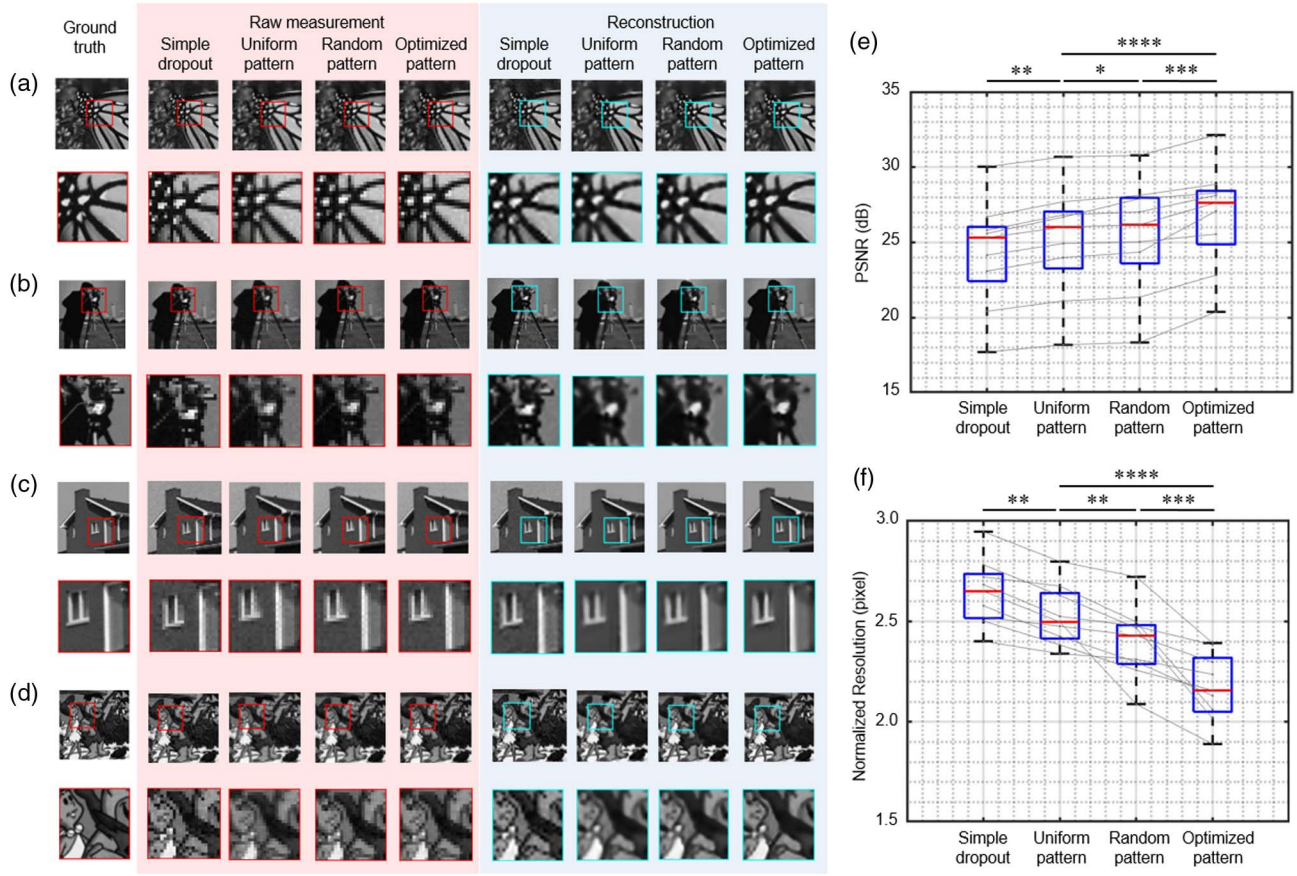

Fig. 6. Comparison of the experimental results using different illumination patterns in the scanning in a CW illumination setting. (a)–(d) Experimental results of the sample: (a) butterfly, (b) cameraman, (c) house, and (d) the Flintstones. The different columns show the ground truth results using high-resolution point scanning, raw measurement using different illumination patterns at an undersampling rate of 6.25%, and the corresponding reconstruction results. (e) PSNR of the reconstructed images for a total of nine samples. (f) Spatial resolution of the reconstructed images for a total of nine samples, calculated from Fourier ring correlation. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$, in one-way ANOVA.

approach. At an undersampling rate of 6.25%, all of the reconstructed results through ISTA-Net show a significant improvement from the raw measurement where high-resolution features are better resolved. Comparing between different illumination patterns, we notice more details on the edges of the reconstructed images in the optimized pattern illumination cases [Figs. 6(a)–6(d)]. Using the metric of PSNR [Fig. 6(e)] and spatial resolution [Fig. 6(f)], we find that DeCIOPS outperforms the simple dropout, the uniform illumination pattern, and the random illumination pattern with an average of 2.41 dB, 1.82 dB, and 1.73 dB improvement in PSNR and an average of 18.9%, 14.0%, and 11.6% improvement on spatial resolution, respectively.

B. Reconstruction Results with a Low-Repetition-Rate Pulsed Light Source

Using the same experimental setup and CW light source, we preprocess the acquired data through a digital undersampling process (Appendix A.5), so the results mimic that acquired by using a low-repetition-rate pulsed light source (Fig. 7). As the integration time of each measured pixel is reduced, the raw measurements have a lower SNR compared to the CW light source setting. Nevertheless, the reconstruction results show a greatly improved quality. On average, the PSNR of the reconstructed images using the optimized illumination pattern shows an improvement of 1.73 dB, 1.41 dB, and 0.64 dB when they are compared with the simple dropout, uniform illumination pattern, and exemplary random illumination pattern, respectively. The improvement in the spatial resolution in the optimized illumination pattern against the simple dropout, the uniform, and the random illumination pattern is 14.06%, 10.33%, and 7.95%, respectively. Compared with the CW light source setting, the performance improvement of the optimized illumination pattern is reduced because of a reduced SNR in the raw measurement in the low-repetition-rate pulsed light source setting. In the Section 5.A, we further discuss how the SNR influences the reconstructed results.

5. DISCUSSION

A. Influence of SNR on Image Reconstruction

In this section, we study the influence of SNR of the raw measurement on the PSNR and pixel resolution of the reconstruction in both simulations and experiments (Fig. 8).

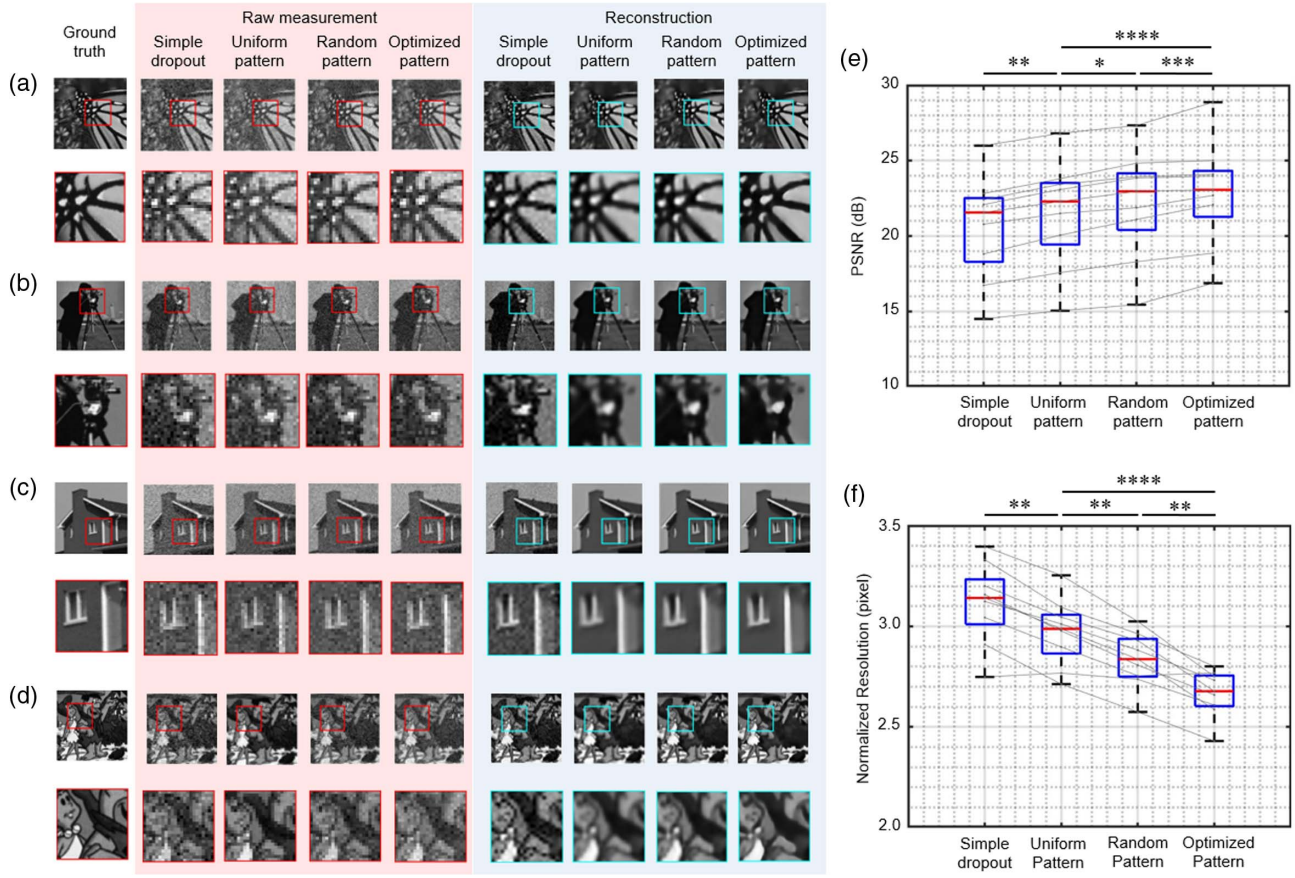

Fig. 7. Comparison of the experimental results using different illumination patterns in the scanning in the low-repetition-rate pulsed light illumination setting. (a)–(d) Experimental results of the sample: (a) butterfly, (b) cameraman, (c) house, and (d) the Flintstones. The different columns show the ground truth results using high-resolution point scanning, raw measurement using different illumination patterns at an undersampling rate of 6.25%, and the corresponding reconstruction results. (e) PSNR of the reconstructed images for a total of nine samples. (f) Spatial resolution of the reconstructed images for a total of nine samples, calculated from Fourier ring correlation. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$, in one-way ANOVA.

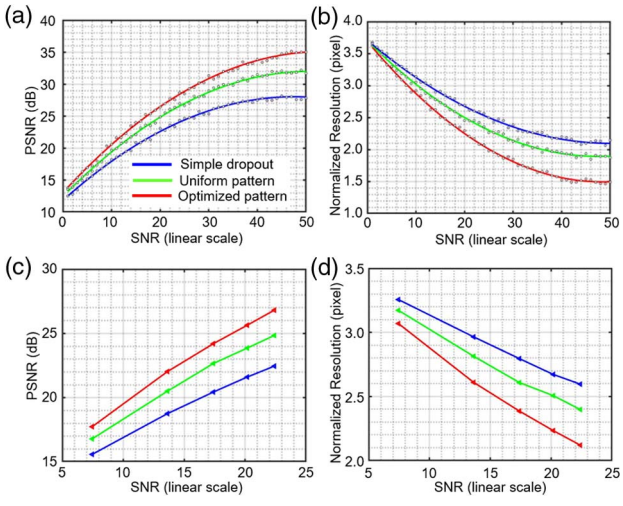

Fig. 8. (a) PSNR and (b) pixel resolution of the reconstructed images versus different SNRs in the raw measurement, for three different sampling patterns (CW configuration), performed through simulation at an undersampling rate of 6.25%. The results were averaged across nine samples used in the experiment and fitted with polynomial curves. (c) and (d) show the experimental results averaged across nine samples.

We performed simulations by adding different noise levels in the measurement. We controlled the SNR in experiments by using different light power. For each SNR level, we trained the DeCIOPS independently. As expected, both the PSNR and pixel resolution of the reconstruction improve as the SNR increases and saturates at a high SNR. Compared with the naïve undersampling through simple dropout (blue curve, Fig. 8) and unweighted average (green curve, Fig. 8), the performance of the optimized illumination pattern (red curve, Fig. 8) has a larger improvement in the PSNR and pixel resolution as SNR increases. This phenomenon emphasizes the advantage of the DeCIOPS, where the optimal encoder and decoder are able to match each other better than other independent untrained encoders. As the noise reduces in the raw measurement (i.e., increasing SNR), the network tends to pay more attention on the reconstruction rather than denoising. As a result, the performance difference among different illumination patterns increases.

B. Compressed Ratio and Size of the Optimized Pattern

In the above simulations and experiments, we set the illumination pattern size such that each pixel in the original

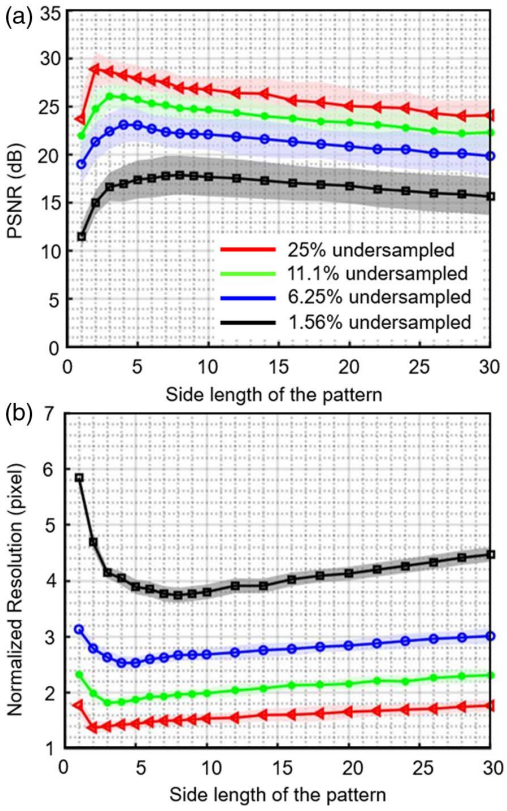

Fig. 9. DeCIOPS reconstruction quality (a) PSNR and (b) pixel resolution dependence on the size of the optimized pattern, for an undersampling rate of 25% (2 × 2, red), 11.1% (3 × 3, green), 6.25% (4 × 4, blue), and 1.5625% (8 × 8, black), across all 79 samples in the validation dataset. Solid curve, mean; shaded area, standard deviation.

high-resolution object is measured only once during the scanning. We term this as the matching condition between the pattern and undersampling rate. Here, for each specific undersampling rate, we simulated the PSNR of the reconstructed images versus the size of the illumination pattern (Fig. 9). We notice that DeCIOPS generates the best reconstruction results when the size of the pattern matches the 2D undersampling rate, where the imaging modality in the encoder is an algebraic linear orthonormal transform. When the size of the pattern is smaller than that of the matching condition, a sharp drop appears in the reconstruction performance. This is attributed to the permanent information loss from the unsampled pixels in between the adjacent measurements, such that the sensing basis Φ^T in Eq. (2) forms a singular transform matrix and thus degrades the quality of reconstruction. When the size of the pattern is larger than that of the matching condition, the sensing basis Φ^T becomes less column orthonormal, resulting in less accurate reconstruction under the CS framework and thus a drop of the PSNR in the reconstructed images.

C. Optimized Pattern versus Random Pattern

In the conventional switching-mask-based single-pixel camera, the random mask is one of the commonly used sampling bases, as it is incoherent with the spatial property of the sample. We found the same in the pattern-scanning scheme: the random pattern shows a superior performance compared with the uniform pattern. However, the optimized pattern, found by the auto-encoder through the end-to-end optimization, outperforms the random pattern. This is expected, as the random pattern attained by the Monte Carlo method is independent of the dataset of a specific task, unlike the optimized pattern. Furthermore, the performance improvement of the optimized pattern over the random pattern increases when the measurements are more highly undersampled, as seen in the comparison between 6.25% (4 × 4 pattern) and 1.5625% (8 × 8 pattern) (Table 1). This is attributed to more trainable parameters in the case of the optimized pattern and more uncertain random variables in the case of the random pattern.

D. Comparison with Conventional Switching-Mask-Based Single-Pixel Camera

As both DeCIOPS and the conventional switching-mask-based single-pixel camera [Fig. 1(b)] leverage compressed sensing, we expect that their reconstruction quality is similar at the same undersampling rate. Indeed, the one-way analysis of variance (ANOVA) test shows no significant difference for PSNR and pixel resolution in the reconstructed images between these two approaches for both simulation and experiment (Fig. 10, Table 2, 256 × 256 pixels high-resolution objects, 6.25% undersampling rate, ISTA-Net reconstruction framework for both DeCIOPS and switching-mask-based single-pixel camera). However, DeCIOPS has a much faster acquisition speed. To image a high-resolution object with 256 × 256 pixels, DeCIOPS acquires 64 × 64 measurements at an undersampling rate of 6.25%. By using a scanning system with an 8 kHz resonant scanner, the acquisition time is 4 ms. In the conventional switching-mask-based single-pixel camera, at the same undersampling rate, it takes 180 ms to cycle $64 \times 64 = 4096$ patterns using a 22.7 kHz high-speed DMD, without considering the integration time of the detector. Thus, the imaging speed of our method is orders of magnitude faster than that of the conventional compressed sensing approach using a single-pixel camera.

Table 1. Comparison of the PSNR and Pixel Resolution across the Uniform Pattern, Random Pattern, and Optimized Pattern between 6.25% and 1.5625% Undersampling Rate in DeCIOPS[a]

Undersampling Rate	PSNR (dB)			Normalized Resolution (pixel)		
	Uniform Pattern	Random Pattern	Optimized Pattern	Uniform Pattern	Random Pattern	Optimized Pattern
6.25%	21.49 ± 1.83	21.82 ± 1.72	23.13 ± 1.54	2.79 ± 0.18	2.68 ± 0.16	2.55 ± 0.15
1.5625%	15.91 ± 2.01	16.78 ± 1.94	18.47 ± 1.75	4.55 ± 0.26	4.28 ± 0.25	3.76 ± 0.21

[a]CW light configuration, SNR ~ 22, for all 79 samples in validation dataset.

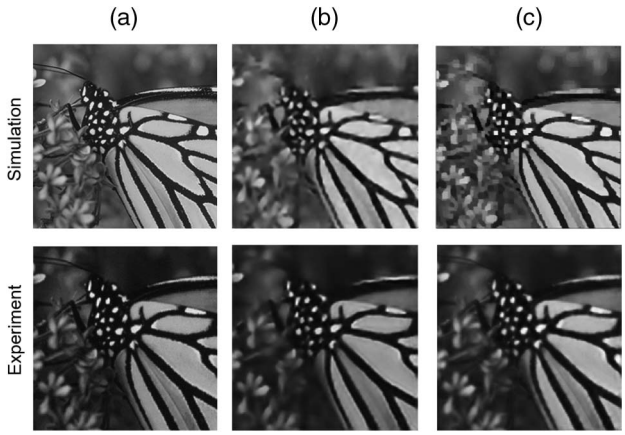

Fig. 10. Comparison of the reconstruction results between DeCIOPS and conventional switching-mask-based single-pixel camera. (a) The ground truth of an original object, butterfly. (b) Reconstruction result of DeCIOPS using ISTA-Net at an undersampling rate of 6.25%. (c) Reconstruction result of the switching-mask-based single-pixel camera imaging approach using ISTA-Net. Top row, simulation. Bottom row, experiment. The ground truth of the experiment is obtained by the high-resolution point scanning.

E. Advantage of ISTA-Net in Image Reconstruction

In DeCIOPS, we apply ISTA-Net as the decoder. ISTA-Net is a CS-induced neural network. Compared with conventional optimization algorithms where the regularization term is designed empirically, ISTA-Net is entirely data-driven and can learn optimized regularization through the neural network. Compared with other neural networks that could be used for super-resolution, such as U-Net [30,31] and densely connected super-resolution network (DCSRN) [32,33], the embedded CS algorithm in ISTA-Net fits better in the motivation of DeCIOPS and other compressed sensing framework, i.e., performing fewer measurements while being able to reconstruct high-resolution images. Indeed, when we compare the PSNR and pixel resolution of the reconstructed images across ISTA-Net, U-Net, and DCSRN, ISTA-Net shows the best performance (Appendix B). In addition to DeCIOPS, we believe ISTA-Net could also benefit other applications such as denoising [34], fast magnetic resonance imaging (MRI) [35], and other super-resolution imaging modalities [18].

F. Advantage of End-to-End Optimized Auto-Encoder and Its Application in Future Imaging Systems

In most existing optical imaging modalities, image formation is empirically designed and optimized, and the deconvolution or object reconstruction algorithm is subsequently tailored to the image formation process. The recent development of low-cost, advanced micro-optics manufacturing techniques, such as three-dimensional (3D) printing and micro–nano-fabrication [36–40], allows rapid prototyping of user-designed optical elements, which opens new opportunities to redesign the image formation process that best fits the specific applications. Instead of sequentially designing the image formation and the reconstruction algorithm, their joint end-to-end optimization produces a global optimal solution [41–43], which is the underlying principle of DeCIOPS. We use an auto-encoder to model the image formation and reconstruction within a single framework and perform end-to-end training to optimize the sampling pattern and ISTA-Net simultaneously. Our results show that the optimized sampling pattern indeed results in the best overall performance. Such an end-to-end training and data-driven approach prevents any empirical bias that may negatively impact the design. We envision that such an approach will enable many challenging applications such as super-resolution imaging [23,42–46], 3D imaging [41,47–49], and high-speed computational cameras [50–52].

G. Applicability in Two-Photon Microscopy

While our imaging system used a CW light source, we mimicked the experimental condition of a pulsed light source and successfully demonstrated the applicability and excellent performance of DeCIOPS. This opens a new avenue to apply DeCIOPS in two-photon microscopy. In conventional two-photon microscopes [5,7,8], the image is acquired through pixel-by-pixel point scanning. While this enables deep tissue imaging as it resists light scattering, it reduces the imaging speed. Recently, there have been multiple reports applying compressed sensing in two-photon microscopy, with the same approach in the conventional switching-mask-based single-pixel camera [53–55]. However, the improvement on the imaging speed is limited due to the low switching speed of DMDs or the liquid-crystal-based spatial light modulators. When applying DeCIOPS in two-photon microscopy, we expect that our approach will significantly increase the imaging speed and will notably benefit functional imaging through two-photon microscopy.

H. Passive Light Illumination

In our experiment, DeCIOPS is implemented using active light illumination (i.e., structured illumination), which is commonly used in biomedical imaging. In other imaging systems, passive light illumination may be preferred. In fact, any passive light illumination wide-field imaging using a focal plane array

Table 2. Comparison of Reconstruction Results between DeCIOPS (CW Light Configuration) and Conventional Switching-Mask-Based Single-Pixel Camera[a,b]

	Simulation		Experiment	
	PSNR (dB)	Normalized Resolution (pixel)	PSNR (dB)	Normalized Resolution (pixel)
DeCIOPS	28.01 ± 1.03	2.13 ± 0.12	27.71 ± 1.18	2.15 ± 0.13
Switching-mask-based single-pixel camera	28.36 ± 1.00	2.12 ± 0.12	27.98 ± 1.16	2.14 ± 0.13

[a]Both simulation and experiment were performed on nine samples from the validation dataset, with an undersampling rate of 6.25% and SNR ~ 22.
[b]No significant difference (one-way ANOVA) is found on the performance metrics between the two imaging modalities.

(i.e., camera) can be converted to DeCIOPS (Appendix C). A scanner can be added to the passive wide-field imaging to scan the entire image originally projected to the camera. By inserting a fixed mask with an appropriate aperture at the plane where the image is scanned, a single-pixel detector can measure the subsampled convolution between the mask and the original image. The object can then be reconstructed using the same algorithms in DeCIOPS for the active light illumination cases. This way, a structured detection version of DeCIOPS can be implemented.

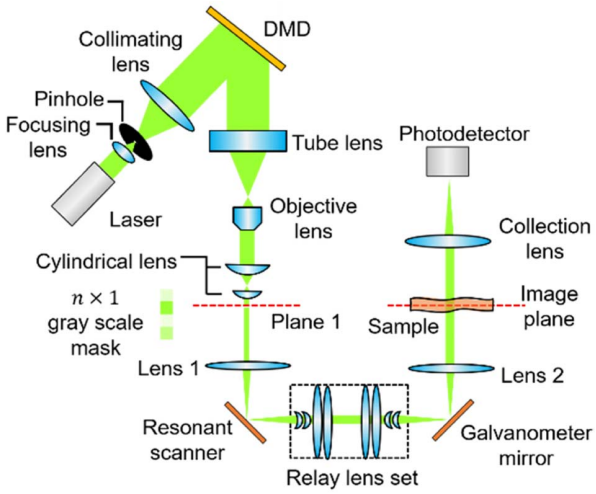

Fig. 11. Experimental setup of DeCIOPS that generates an $n \times 1$ size pattern and scans it across the sample. The setup is similar to that generating the $n \times n$ size pattern shown in Fig. 5, but with a $4f$ system composed of cylindrical lenses added after the objective lens to shrink the original $n \times n$ size pattern in one dimension by a factor of n into the $n \times 1$ size. The red dashed line (plane 1 and object plane) indicates the conjugate plane of the gray-scale pattern mask.

6. CONCLUSION

We demonstrated a new high-speed imaging modality, DeCIOPS, by synthesizing the strength of conventional point scanning and single-pixel camera through compressed sensing. The high-speed imaging arises from the fast beam-scanning mechanism and a highly efficient sampling scheme through compressed sensing; meanwhile, an auto-encoder framework allows the simultaneous optimization of the image formation and reconstruction process in DeCIOPS. We validated DeCIOPS through both simulation and experiments in both CW and pulsed light source conditions. This new image modality can be adapted to any existing imaging systems using beam scanning, such as confocal microscope and two-photon microscope, or wide-field cameras with an added scanning system, and will benefit broad applications requiring high-speed imaging.

APPENDIX A: EXPERIMENTAL SETUP OF DECIOPS

1. Optical Setup

The optical setup of DeCIOPS illustrated in Fig. 5 scans an $n \times n$ pattern on the object and is suitable for low-repetition-rate pulsed light illumination settings. For CW light or high-repetition-rate pulsed light illumination, an $n \times 1$ pattern is scanned. When this pattern sweeps n columns, the information of $n \times n$ pixels can be integrated into a single measurement. To generate the $n \times 1$ pattern, we set the column to be identical in the $n \times n$ pattern and add a $4f$ system composed of cylindrical lenses after the objective lens. The $n \times n$ pattern is then shrunk in one dimension by a factor of n into the $n \times 1$ pattern (Fig. 11). The parameters of the lenses used in the setup are listed in Table 3.

2. Pattern Generation

The pattern is generated from a DMD illuminated by a parallel beam. In our experiment, the pattern has 4×4 pixels, each of

Table 3. Detailed Parameters of the Optical Components Used in the Imaging System

Element	Manufacturer	Part Number	Note
Focusing lens	Thorlabs	A397TM-A	Aspherical lens, focal length 11 mm
Pinhole	Thorlabs	P30D	30 μm pinhole
Collimating lens	Thorlabs	AC254-100-AB-ML	Achromatic lens, focal length 100 mm (for pattern scanning)
	Thorlabs	ACL5040-A	Aspherical lens, focal length 40 mm (for point scanning)
Tube lens	Thorlabs	SM2V10	Focal length 200 mm
Objective lens	Olympus	RMS4X	4× objective lens
Cylindrical lens	Thorlabs	LJ1014L1-A	Focal length 25.4 mm
	Thorlabs	LJ1227L2-A	Focal length 6.35 mm
Lenses 1 and 2	OptoSigma	SLB-60-250 P	Focal length 250 mm
Relay lens set [56] (one set listed	Thorlabs	LC1582-A	Focal length −75 mm
here; two sets arranged	Thorlabs	LC1582-A	Focal length −75 mm
symmetrically are required in	Thorlabs	LE1076-A	Meniscus lens, focal length 100 mm
setup)	Thorlabs	LA1399-A	Focal length 175 mm
	Thorlabs	LA1050-A	Focal length 100 mm
	Thorlabs	LA1727-A	Focal length 750 mm
Collection lens	Thorlabs	AC508-075-A-ML	Achromatic lens, focal length 75 mm
Galvanometer scanner	Cambridge Technology	6215HM40B	
Resonant scanner	Cambridge Technology	CRS 8 kHz	Resonant frequency 8 kHz
DMD	Texas Instruments	DLP7000	

which has a gray-scale throughput/intensity ranging from 0 to 1. As each DMD mirror pixel only provides a binary (on or off) light throughput, we group 32 × 32 DMD mirror pixels as one super-pixel, so a gray-scale light throughput becomes feasible within one super-pixel. Each super-pixel thus provides 1025 states in the range from 0 to 1. The multistate light throughput is achieved by randomly selecting parts of the mirrors to be on, the number of which is the product of the desired light throughput and total number of mirror pixels (1024).

In the case of 4 × 1 pattern scanning, another $4f$ system composed of cylindrical lenses is added after the objective lens (Fig. 11). Such a system shrinks the 4 × 4 pattern in one dimension by a factor of four.

To obtain the high-resolution ground truth image, we reconfigure the imaging system for point scanning. We replace the collimating lens between the pinhole and the DMD to the one with shorter focal length and set the displayed pattern on the DMD to be uniform. This reduces the beam size on the image plane, and the pattern effectively turns into a spot matching the size of a single super-pixel of the 4 × 4 pattern. This ensures the same overall energy collected by the photodetector and thus a similar SNR per measurement as in DeCIOPS.

3. Characterization of the Pattern

As the optical mode from the diode laser was not symmetric, we implemented a spatial filter using a pinhole after the diode laser output. Nevertheless, non-uniform intensity of the light illuminated on the DMD still appeared. Furthermore, the laser interference could corrupt the pattern. We, thus, fine-tuned the

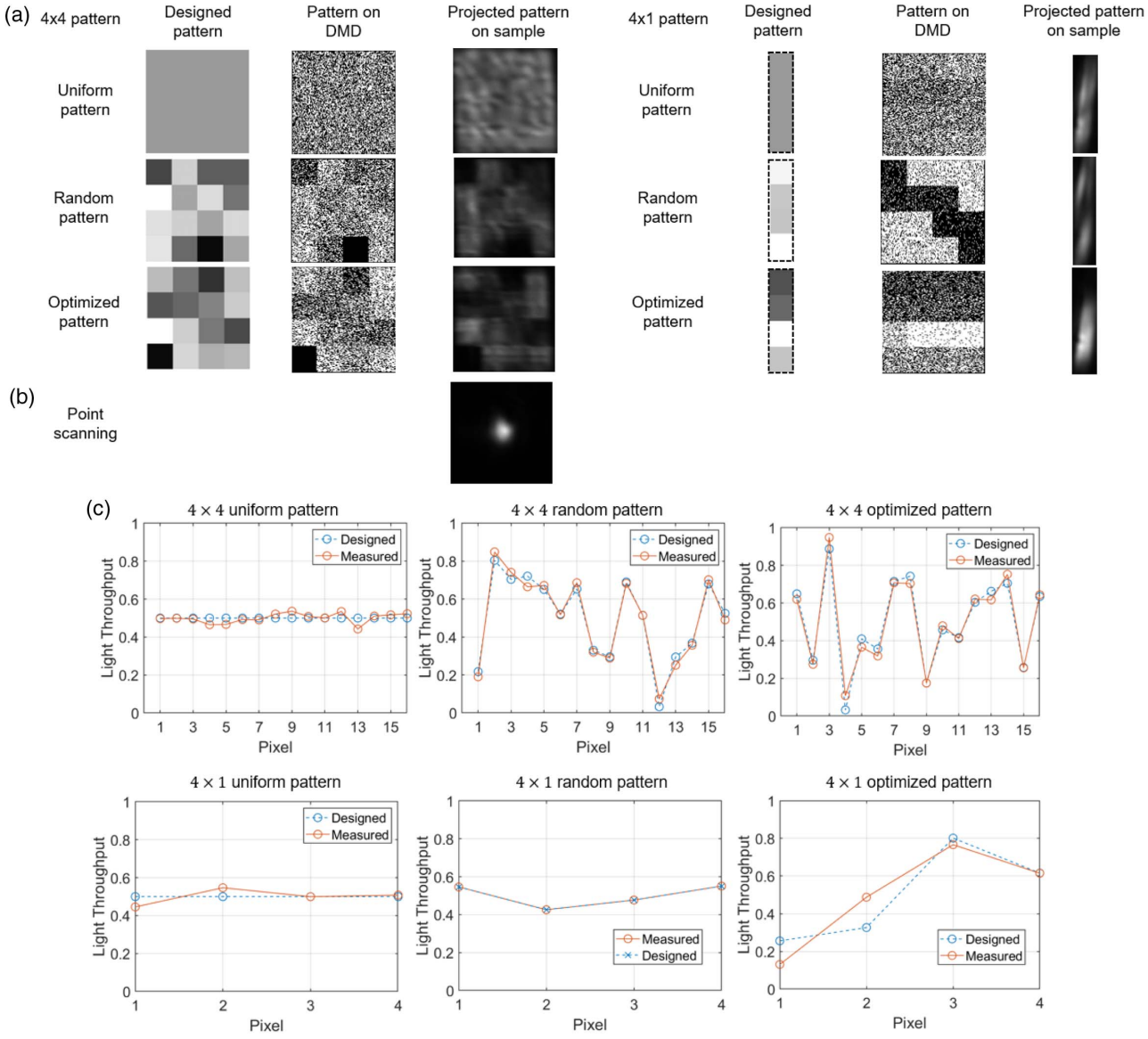

Fig. 12. (a) Measured patterns on the sample (super-pixels) match well with the designed patterns. Each gray-scale super-pixel is generated by 32 × 32 binary pixels in the DMD. The left panel shows the cases for 4 × 4 patterns, and the right panel shows the cases for the 4 × 1 pattern. (b) A single spot pattern is generated for conventional point-scanning imaging to obtain the high-resolution ground truth of the sample. The spot size matches the size of a super-pixel. (c) Pixel-by-pixel comparisons between the measured patterns on the sample and the designed patterns show excellent matchings between the two.

DMD pixel value to calibrate the intensity of the super-pixels, so their values were as close to the design as possible.

We used an iterative approach to calibrate the intensity of the super-pixel. The design pattern M_0 with the gray-scale super-pixel, which was also written as $M^{(0)}$ for consistency, was first converted to a pattern $W^{(0)}$ on the DMD through an operator $\mathcal{D}$. $\mathcal{D}$ essentially converted each super-pixel in M_0 into 32×32 pixels in binary values on the DMD, with the on and off pixels randomly distributed. $\mathcal{D}^{-1}$ is the inverse operation. We used a camera to capture the projected pattern $M^{(1)}$ on the object plane. We then compared $M^{(1)}$ and M_0 and updated $W^{(0)}$ into $W^{(1)}$ through the following algorithm:

$$W^{(1)} = \mathcal{D}((M^{(1)} - M_0) \times a + M^{(0)}), \quad \textbf{(A1)}$$

where a is the step size to control the update rate. For the kth iteration, we have

$$W^{(k)} = \mathcal{D}((M^{(k)} - M_0) \times a + \mathcal{D}^{-1}(W^{(k-1)})). \quad \textbf{(A2)}$$

In general, the pattern could be calibrated well with <5 iterations. With this calibration, we verified that the projected pattern on the sample matched the designed pattern in most of the cases (Fig. 12). In the special case of the 4×1 random pattern, we took an alternative approach. We randomly generated the pattern and measured it at the object plane. We then used this pattern as the "design" to train the decoder for object reconstruction.

4. Pixel Resolution and Field of View

The size of an individual DMD mirror is 13.6 μm × 13.6 μm, and thus the size of a super-pixel is $(32 \times 13.6) \times (32 \times 13.6) = 435.2$ μm × 435.2 μm. The imaging system has a magnification of 4.44, resulting in a size of $\left(\frac{435.2}{4.44}\right) \times \left(\frac{435.2}{4.44}\right) =$ 97.9 μm × 97.9 μm of a super-pixel on the image plane. A full resolution object with 256×256 pixels (~24.5 mm × 24.5 mm), requires a scanning range of $\pm 2.8° \times \pm 2.8°$ in the scanner set. We measured the illumination pattern on the object plane across different scanning angles (Fig. 13). The excellent field uniformity ensures the quality of DeCIOPS.

5. Data Acquisition

Data acquisition is performed using a high-speed data acquisition card vDAQ and ScanImage software (Vidrio Technologies). As the illumination pattern continuously scans across each row, the data acquisition card samples the data from the photodetector at a rate higher than the single-pixel rate. The data acquired within the duration of a single pixel is then automatically averaged/integrated and saved as a single-pixel value. Compared with the full resolution single-point-scanning condition (256×256 pixels), we reduce the single-pixel rate and the number of scanning lines by 75% in DeCIOPS (CW light setting), reaching an undersampling rate of 6.25% (64×64 pixels of measurement).

To mimic the low-repetition-rate pulsed light source condition, we sample the object with a high resolution at 2048×2048 pixels (corresponding to a high single-pixel rate) and then downsample the acquired image digitally into 64×64 pixels by dropping all of the other pixels. Here, each pixel has a small average/integration duration and could thus be considered as being acquired by a single light pulse.

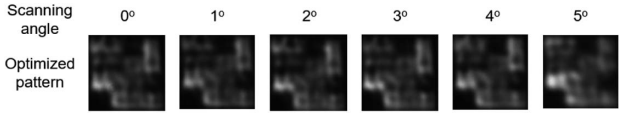

Fig. 13. Measured patterns at the image plane stay consistent across different scanning angles.

6. Estimation of Signal-to-Noise Ratio

To calculate the SNR of the image in the experiment, we acquire the same image 20 times. For each pixel, we calculate the signal μ and the noise σ as the mean and the standard deviation across 20 measurements, respectively. The SNR for the pixel is then estimated as μ/σ. The SNR of the entire image is taken as the average of the SNR of all pixels.

APPENDIX B: COMPARISON BETWEEN ISTA-NET, U-NET, AND DCSRN

We compare the performance of ISTA-Net, U-Net, and DCSRN in object reconstruction. In the auto-encoder framework, the decoder implemented by ISTA-Net is replaced by U-Net or DCSRN. U-Net is widely used in image reconstruction and segmentation. It first condenses the size of the input images to extract its context and feature and then grows them in an expanding path to perform local reconstruction [30]. DCSRN, derived from densely connected convolutional networks [33], has a faster training speed and accurate reconstruction results and is commonly used in applications such as 2D or 3D biomedical super-resolution imaging. In addition to ISTA-Net, U-Net, and DCSRN, we used B-spline interpolation [28] to reconstruct the object undersampled through a simple dropout as a baseline. All of the simulation was completed on GPU RTX1080Ti 11 GB with 200 epochs and a batch size of five. In each decoder except for the B-spline, we learned an optimized illumination pattern. We used the validation data set to evaluate the PSNR and pixel resolution of the reconstructed objects. U-Net, DCSRN, and ISTA-Net all outperform B-spline interpolation. While U-Net and DCSRN do not show a significant difference in performance, ISTA-Net outperforms both U-Net and DCSRN with a ~1.85 dB increase in PSNR and 6.55% improvement in resolution at an undersampling rate of 6.25% (Fig. 14). The simulation results demonstrated a clear advantage of ISTA-Net, which is a CS inspired neural network, in DeCIOPS.

APPENDIX C: DECIOPS USING PASSIVE LIGHT ILLUMINATION

To implement DeCIOPS in a passive light illumination setting (i.e., structured detection), which is commonly used in photography, a scanner can be added to the passive wide-field imaging system to scan the entire image originally projected to the camera. By inserting a fixed mask with an appropriate aperture at the plane where the image is scanned, a single-pixel detector can measure the subsampled convolution between the mask and the original image (Fig. 15). The object can then be

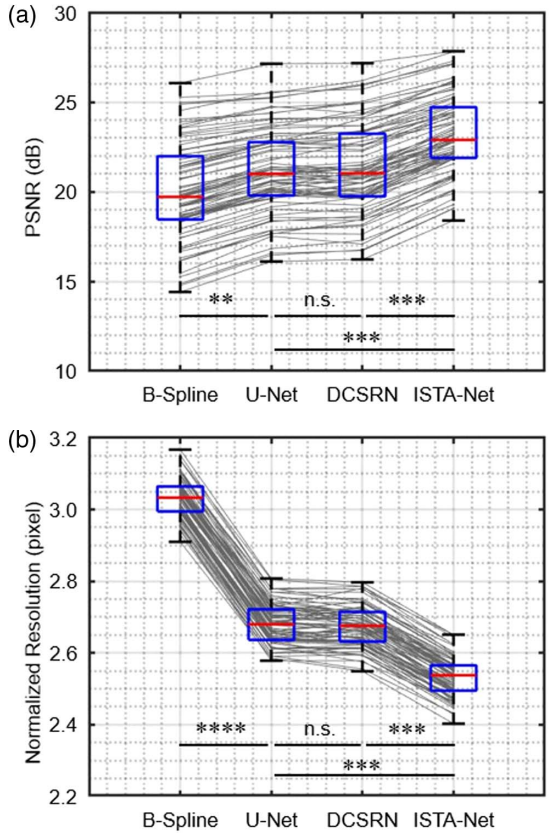

Fig. 14. Comparison of (a) PSNR and (b) pixel resolution of the reconstructed objects of all 79 samples in the validation dataset for B-spline, U-Net, DCSRN, and ISTA-Net in the auto-encoder framework, at an undersampling rate of 6.25%. n.s., not significant; **, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$, in one-way ANOVA.

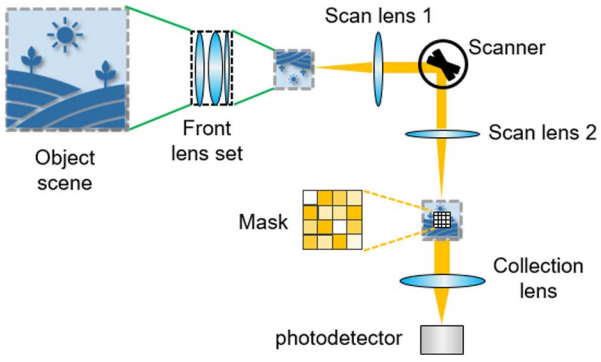

Fig. 15. Optical setup of DeCIOPS with passive light illumination (i.e., structured detection) for applications such as photography.

reconstructed using the same algorithms in DeCIOPS for the active light illumination cases.

Funding. National Science Foundation (CARRER 1847141); National Institute of Neurological Disorders and Stroke (R01NS118289); National Eye Institute (R01NS118289); Burroughs Wellcome Fund (CASI 1015761).

Disclosures. The authors declare no conflicts of interest.

REFERENCES

1. A. Rogalski, *Infrared Detectors* (CRC Press, 2010).
2. 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**, 83–91 (2008).
3. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," Nat. Photonics **13**, 13–20 (2019).
4. G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," Opt. Express **28**, 28190–28208 (2020).
5. W. Denk, J. H. Strickler, and W. W. Webb, "Two-photon laser scanning fluorescence microscopy," Science **248**, 73–76 (1990).
6. J. B. Pawley, *Handbook of Biological Confocal Microscopy* (Springer, 2006).
7. W. R. Zipfel, R. M. Williams, and W. W. Webb, "Nonlinear magic: multiphoton microscopy in the biosciences," Nat. Biotechnol. **21**, 1369–1377 (2003).
8. F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," Nat. Methods **2**, 932–940 (2005).
9. L. Beiser, "Fundamental architecture of optical scanning systems," Appl. Opt. **34**, 7307–7317 (1995).
10. E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians* (2006), pp. 1433–1452.
11. E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," Inverse Prob. **23**, 969 (2007).
12. E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," IEEE Signal Process. Mag. **25**, 21–30 (2008).
13. A. P. Spencer, B. Spokoyny, S. Ray, F. Sarvari, and E. Harel, "Mapping multidimensional electronic structure and ultrafast dynamics with single-element detection and compressive sensing," Nat. Commun. **7**, 10434 (2016).
14. L. Bian, J. Suo, Q. Dai, and F. Chen, "Experimental comparison of single-pixel imaging algorithms," J. Opt. Soc. Am. A **35**, 78–87 (2018).
15. J. B. Sampsell, "Digital micromirror device and its application to projection displays," J. Vac. Sci. Technol. B **12**, 3242–3246 (1994).
16. Z. H. Xu, W. Chen, J. Penuelas, M. Padgett, and M. J. Sun, "1000 fps computational ghost imaging using LED-based structured illumination," Opt. Express **26**, 2427–2434 (2018).
17. H. Wu, Z. Zheng, Y. Li, W. Dai, and H. Xiong, "Compressed sensing via a deep convolutional auto-encoder," in *IEEE Visual Communications and Image Processing (VCIP)* (2018), pp. 1–4.
18. J. Zhang and B. Ghanem, "ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing," in *IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 1828–1837.
19. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imag. Sci. **2**, 183–202 (2009).
20. N. Pavillon and N. I. Smith, "Compressed sensing laser scanning microscopy," Opt. Express **24**, 30038–30052 (2016).
21. L. Fang, F. Monroe, S. W. Novak, L. Kirk, C. R. Schiavon, S. B. Yu, T. Zhang, M. Wu, K. Kastner, Y. Kubota, Z. Zhang, G. Pekkurnaz, J. Mendenhall, K. Harris, J. Howard, and U. Manor, "Deep learning-based point-scanning super-resolution imaging," bioRxiv: 740548 (2019).
22. B. Li, Q.-R. Yan, Y.-F. Wang, Y.-B. Yang, and Y.-H. Wang, "A binary sampling Res2net reconstruction network for single-pixel imaging," Rev. Sci. Instrum. **91**, 033709 (2020).
23. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Gunaydin, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nat. Methods **16**, 103–110 (2019).
24. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255.
25. K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok, "Reconnet: non-iterative reconstruction of images from compressively sensed measurements," in *IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 449–458.

26. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *8th IEEE International Conference on Computer Vision (ICCV)* (2001), pp. 416–423.
27. S. Koho, G. Tortarolo, M. Castello, T. Deguchi, A. Diaspro, and G. Vicidomini, "Fourier ring correlation simplifies image restoration in fluorescence microscopy," Nat. Commun. **10**, 3103 (2019).
28. T. M. Lehmann, C. Gonner, and K. Spitzer, "Addendum: B-spline interpolation in medical image processing," IEEE Trans. Med. Imaging **20**, 660–665 (2001).
29. T. A. Pologruto, B. L. Sabatini, and K. Svoboda, "ScanImage: flexible software for operating laser scanning microscopes," Biomed. Eng. Online **2**, 13 (2003).
30. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (2015), pp. 234–241.
31. J. Feng, J. Deng, Z. Li, Z. Sun, H. Dou, and K. Jia, "End-to-end Res-Unet based reconstruction algorithm for photoacoustic imaging," Biomed. Opt. Express **11**, 5321–5340 (2020).
32. Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, "Brain MRI super resolution using 3D deep densely connected neural networks," in *IEEE 15th International Symposium on Biomedical Imaging (ISBI)* (2018), pp. 739–742.
33. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 4700–4708.
34. B. Lee, B. Ku, W.-J. Kim, S. Kim, and H. Ko, "Denoising ISTA-Net: learning based compressive sensing with reinforced non-linearity for side scan sonar image denoising," J. Acoust. Soc. Korea **39**, 246–254 (2020).
35. E. D. W. N. Pezzotti, S. Yousefi, M. S. Elmahdy, J. van Gemert, C. Schülke, M. Doneva, T. Nielsen, S. Kastryulin, B. P. F. Lelieveldt, M. J. P. van Osch, and M. Staring, "Adaptive-CS-Net: FastMRI with adaptive intelligence," arXiv:1912.12259 (2019).
36. X. Q. Zhou, Y. H. Hou, and J. Q. Lin, "A review on the processing accuracy of two-photon polymerization," AIP Adv. **5**, 030701 (2015).
37. T. Gissibl, S. Thiele, A. Herkommer, and H. Giessen, "Two-photon direct laser writing of ultracompact multi-lens objectives," Nat. Photonics **10**, 554–560 (2016).
38. J. W. Li, P. Fejes, D. Lorenser, B. C. Quirk, P. B. Noble, R. W. Kirk, A. Orth, F. M. Wood, B. C. Gibson, D. D. Sampson, and R. A. McLaughlin, "Two-photon polymerisation 3D printed freeform micro-optics for optical coherence tomography fibre probes," Sci. Rep. **8**, 14789 (2018).
39. N. Mohammad, M. Meem, X. Wan, and R. Menon, "Full-color, large area, transmissive holograms enabled by multi-level diffractive optics," Sci. Rep. **7**, 5789 (2017).
40. N. Vaidya and O. Solgaard, "3D printed optics with nanometer scale surface roughness," Microsys. Nanoeng. **4**, 18 (2018).
41. Y. C. Wu, V. Boominathan, H. J. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "PhaseCam3D-learning phase masks for passive single view depth estimation," in *IEEE International Conference on Computational Photography* (2019), pp. 1–12.
42. V. Sitzmann, S. Diamond, Y. F. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," ACM Trans. Graph. **37**, 114 (2018).
43. Q. L. Sun, J. Zhang, X. Dun, B. Ghanem, Y. F. Peng, and W. Heidrich, "End-to-end learned, optically coded super-resolution SPAD camera," ACM Trans. Graph. **39**, 9 (2020).
44. Y. Li, B. Sixou, and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems," IRBM (2020).
45. W. Ouyang, A. Aristov, M. Lelek, X. Hao, and C. Zimmer, "Deep learning massively accelerates super-resolution localization microscopy," Nat. Biotechnol. **36**, 460–468 (2018).
46. M. Weigert, U. Schmidt, T. Boothe, A. Muller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nat. Methods **15**, 1090–1097 (2018).
47. S. Zhang, "High-speed 3D shape measurement with structured light methods: a review," Opt. Laser Eng. **106**, 119–131 (2018).
48. Y. C. Wu, Y. Rivenson, H. D. Wang, Y. L. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," Nat. Methods **16**, 1323–1331 (2019).
49. H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, "Deep learning for single-shot autofocus microscopy," Optica **6**, 794–797 (2019).
50. Y. Ma, Y. Lee, C. Best-Popescu, and L. Gao, "High-speed compressed-sensing fluorescence lifetime imaging microscopy of live cells," Proc. Natl. Acad. Sci. USA **118**, e2004176118 (2020).
51. D. Qi, S. Zhang, C. Yang, Y. He, F. Cao, J. Yao, P. Ding, L. Gao, T. Jia, and J. Liang, "Single-shot compressed ultrafast photography: a review," Adv. Photon. **2**, 014003 (2020).
52. L. Gao, J. Y. Liang, C. Y. Li, and L. H. V. Wang, "Single-shot compressed ultrafast photography at one hundred billion frames per second," Nature **516**, 74–77 (2014).
53. M. Alemohammad, J. Shin, D. N. Tran, J. R. Stroud, S. P. Chin, T. D. Tran, and M. A. Foster, "Widefield compressive multiphoton microscopy," Opt. Lett. **43**, 2989–2992 (2018).
54. P. Wijesinghe, A. Escobet-Montalban, M. Z. Chen, P. R. T. Munro, and K. Dholakia, "Optimal compressive multiphoton imaging at depth using single-pixel detection," Opt. Lett. **44**, 4981–4984 (2019).
55. Y.-X. Ren, C. Kong, H. He, X. Zeng, K. K. Tsia, and K. K. Wong, "Encrypted wide-field two-photon microscopy with single-pixel detection and compressed sensing," Appl. Phys. Express **13**, 032007 (2020).
56. J. N. Stirman, I. T. Smith, M. W. Kudenov, and S. L. Smith, "Wide field-of-view, multi-region, two-photon imaging of neuronal activity in the mammalian brain," Nat. Biotechnol. **34**, 857–862 (2016).