

RESEARCH ARTICLE | MAY 04 2023

Photon-starved snapshot holography

Yunping Zhang ; Stanley H. Chan ; Edmund Y. Lam 

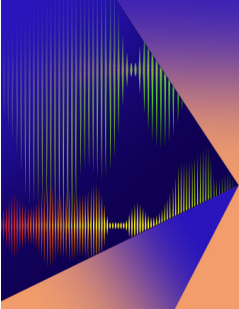


APL Photonics 8, 056106 (2023)

<https://doi.org/10.1063/5.0145833>



CrossMark



APL Photonics

Special Topic: State-of-the-art and Future Directions in Optical Frequency Combs

Submit Today!

Photon-starved snapshot holography

Cite as: APL Photon. 8, 056106 (2023); doi: 10.1063/5.0145833

Submitted: 8 February 2023 • Accepted: 13 April 2023 •

Published Online: 4 May 2023



Yunping Zhang,¹  Stanley H. Chan,²  and Edmund Y. Lam^{1,a)} 

AFFILIATIONS

¹ Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

² School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana 47907, USA

^{a)} Author to whom correspondence should be addressed: elam@eee.hku.hk

ABSTRACT

Digital holography (DH) is a powerful imaging modality that is capable of capturing the object wavefront information, making it very valuable for diverse scientific research applications. Generally, it requires ample illumination to enable good fringe visibility and a sufficient signal-to-noise ratio. As such, in situations such as probing live cells with minimal light interaction and high-speed volumetric tracking in flow cytometry, the holograms generated with a limited photon budget suffer from poor pattern visibility. While it is possible to make use of photon-counting detectors to improve the hologram quality, the long recording procedure coupled with the need for mechanical scanning means that real-time extremely low-light holographic imaging remains a formidable challenge. Here, we develop a snapshot DH that can operate at an ultra-low photon level (less than one photon per pixel). This is achieved by leveraging a quanta image sensor to capture a stack of binary holographic frames and then computationally reconstructing the wavefront through integrating the mathematical imaging model and the data-driven processing, an approach that we termed PSHoloNet. The robustness and versatility of our DH system are demonstrated on both synthetic and experimental holograms with two common DH tasks, namely particle volumetric reconstruction and phase imaging. Our results demonstrate that it is possible to expand DH to the photon-starved regime, and our method will enable more advanced holography applications in various scientific imaging systems.

© 2023 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0145833>

I. INTRODUCTION

Digital holography (DH) is a versatile optical imaging technique that records an interference pattern, known as a hologram, on an electronic sensor, which can then be computationally processed to retrieve wavefront information.¹ Because of this unique capability, it is used in a wide range of applications, such as biomedical imaging,^{2,3} environmental research,^{4,5} and particle field reconstruction.^{6,7} Typically, good illumination and a sufficient exposure time are needed to provide clear holographic patterns; such requirements unfortunately restrict the further use of this powerful modality under more challenging imaging conditions. For instance, for some biological samples or photo-materials vulnerable to light irradiation, there is strict control of the illumination power in consideration of the phototoxicity of the specimen.⁸ Similarly, for high-speed particle tracking, the exposure time has to be very short in order to capture fast dynamics.⁹ There is a clear need to detect such weak signals amidst a noise level that is possibly

significantly higher.¹⁰ In DH, we can observe in Fig. 1 that the low signal-to-noise ratio (SNR) causes the hologram quality to deteriorate substantially, thus posing great challenges for conventional reconstruction algorithms to extract meaningful information from the holograms containing highly contaminated fringe patterns.

These situations necessitate research in *photon-starved holography* since there are scarce photons to form conventional holographic patterns at the sensor plane. Various approaches have been developed to enable DH at a low signal power.^{11–13} Without introducing extra hardware, the purely computational methods are mostly based on noise removal and signal enhancement, such as angular matching¹² and deep-learning transformation.¹³ However, conventional image sensors, such as CCD and CMOS image sensors used for registering holograms have a minimum photon flux requirement of around 10^4 photons per pixel (PPP).¹⁴ As one pushes the capability of DH to extremely low photon flux situations, i.e., one PPP or lower, a new hologram formation mechanism with single-photon detection has to be considered.

Yamamoto *et al.* first attempted to tackle this challenging problem by using a photon-counting detector (PCD).¹⁵ Benefiting from the continuous progress in sensor development, Demoli *et al.*¹⁶ then used a single-photon avalanche photodiode (SPAD) to record the holographic information of signals overshadowed by noise. More recently, Kang *et al.* proposed a deep neural network for retrieving phase information at low photon counts using random phase modulation.¹⁷ In these methods, the long recording time for the mechanical scanning process (as long as 30 h¹⁶) and the extra hardware for phase generation¹⁷ restrict their practical deployment. Therefore, we further aim to make the photon-starved holography a *snapshot imaging system*, which can enable the *in vitro* capture of live cells with minimal sample interaction and high-speed volumetric imaging.

To make photon-starved snapshot holography possible, an image detector with photon-level sensitivity and high resolution is necessary. Among various photon-counting image sensors reported over the past decade,^{18–22} the quanta image sensor (QIS) stands out with its small pixel size (e.g., $\leq 1 \mu\text{m}$), high photon detection efficiency, and ultra-low readout noise. This leads to various applications in the low quanta flux regimes, including security, night vision, and quantum computing.^{23–25}

As reported in Ref. 26, a QIS can be realized using a high-gain photodiode based on existing CMOS techniques^{18–20} or using a SPAD.^{21,22} In this paper, we primarily focus on the CMOS-based QIS although our mathematical formulation can be extended to a SPAD-based QIS. Among the many properties of these sensors, we consider the one-bit mode, where the measured signal is either 1 or 0. This would allow the fastest sampling rate for moving scenes. Yet, integrating this photodetector in DH is not a trivial task, due to its binary signal detection mechanism and very different noise characteristics, which would require specifically designed reconstruction algorithms.

In this work, we overcome several difficulties to present a photon-starved snapshot holography system, which integrates a QIS to take advantage of its low-light detection, and a physics-informed wavefront reconstruction network called PSHoloNet that we have

designed to process the captured signals. Through numerical simulations, we demonstrate the generalization capability of the proposed PSHoloNet from a simplified sensor model during training to a more advanced sensor model during inference. To our best knowledge, it is the first time that a QIS is integrated into a coherent imaging system. To demonstrate the feasibility through synthetic experiments, we show two DH applications, namely 3D particle volumetric reconstruction and phase imaging.

II. METHOD

A. System design and image forward modeling

The Gabor holographic imaging setup together with a QIS as the recording device is designed and shown in Fig. 2. The object $\vartheta(x, y) = A(x, y)e^{j\phi(x, y)}$ is placed at the axial position $z = 0$, with A and ϕ representing its transmittance and phase response, respectively. Suppose that we have an incident plane wave with wavelength λ . Based on the diffraction theory,²⁷ the transmitted wave received at the sensor plane is

$$O(x, y; z) = h_z(x, y) \otimes \vartheta(x, y), \quad (1)$$

where $\otimes$ stands for the convolution and h_z denotes the impulse response of free-space propagation over a distance z . It is represented using the Fresnel approximation as

$$h_z(x, y) = \frac{1}{j\lambda z} e^{j\frac{2\pi}{\lambda} z} e^{j\frac{\pi}{\lambda z}(x^2 + y^2)}. \quad (2)$$

For a 3D volume that is discretized as multiple slices $[\vartheta_1, \vartheta_2, \dots, \vartheta_L]$, assuming that there is no inter-object interference and any nonlinear terms are treated as noise (e.g., twin image and the DC term), the intensity pattern (hologram) captured by the image sensor is

$$I(x, y) = \frac{1}{L} \left\| \sum_{\ell=1}^L O_{\ell}(x, y; z) \right\|^2 = \frac{1}{L} \left\| \sum_{\ell=1}^L h_{z_{\ell}} \otimes \vartheta_{\ell} \right\|^2. \quad (3)$$

Supposing that the hologram created at the sensor plane is discretized by the M unit space area, where the light exposure remains constant over an exposure time (M degree of freedom). It can be represented by a 1D vector (i.e., $\mathbf{I} = [i_1, i_2, \dots, i_M]$) through raster scanning, and each element is expressed using the matrix notation for a compact form as

$$i_m = [\mathbf{H}\mathbf{v}]_m, \quad m \in [1, M], \quad (4)$$

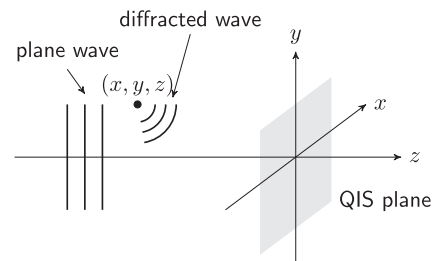


FIG. 2. Illustration of the proposed imaging system.

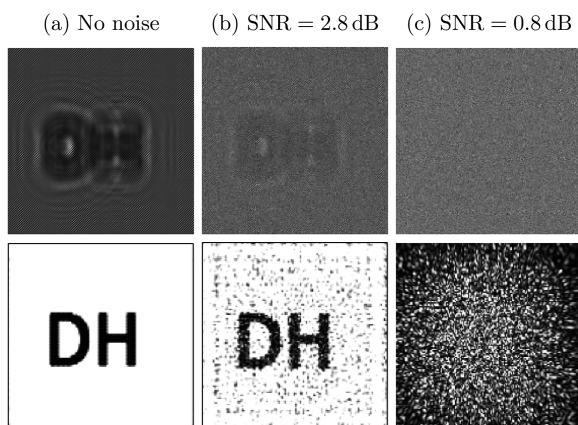


FIG. 1. Reconstruction results from holograms under different SNRs: (a) ideal condition without noise, (b) SNR = 2.8 dB, and (c) SNR = 0.8 dB. Upper row: Holograms recorded under different noise levels. Bottom row: Reconstruction results using backward propagation.

where $\mathbf{v} \in \mathbb{C}^Q$ is a vector corresponding to the stack of discretized object light response in the 3D volume with $Q = LM$, each column of $\mathbf{H} \in \mathbb{C}^{M \times Q}$ is a discretized representation of the impulse response kernel h_{z_ℓ} defined in Eq. (2) at distance z_ℓ , and $[\cdot]_m$ denotes the m -th element of a vector.

The QIS, which we use to record the hologram, is composed of an array of specialized photodetectors called jots.²⁰ Because of the small jot size and fast temporal response, a QIS is regarded as an oversampling device in both spatial and temporal dimensions with sampling ratios K_s and K_T , respectively.^{28,29} We use the overall upsampling factor $K = K_s K_T$ for simplicity of notation. Therefore, the output of a QIS can be considered as a spatial-temporal cube with $N = KM$ units. For a set of units responding to the m -th discrete hologram area with light exposure i_m , the accumulated light flux at a single unit is

$$\varphi_n = \frac{i_m}{K} = \frac{[\mathbf{H}\mathbf{v}]_m}{K}, \quad \text{for } mK \leq n < (m+1)K. \quad (5)$$

Depending on the local values of φ_n , each jot collects a number of photons impinging on its surface, which is represented by Y_n . The relationship between φ_n and photon count Y_n is stochastic and is described by a Poisson-Gaussian distribution,²⁸

$$Y_n = \text{Poisson}(\varphi_n + \eta_{\text{dark}}) + \mathcal{N}(0, \sigma^2), \quad (6)$$

where η_{dark} is the dark current and σ is the standard deviation of the read noise. Since the dark noise of a QIS is negligible,³⁰ the probability distribution of Y_n is

$$p_{Y_n}(y) = \sum_{k=0}^{\infty} \frac{\varphi_n^k e^{-\varphi_n}}{k!} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-k)^2}{2\sigma^2}}. \quad (7)$$

In a photosensitive device, the accumulated photons are converted to electrical signals through a quantization process. In a traditional sensor, the digital signals are generated from an A/D converter and composed of 8–14 bits. On the contrary, a QIS follows a binary (i.e., 1-bit) quantization scheme. Let q be the threshold, and the quantized binary output B_n at each unit is, then,

$$B_n = \begin{cases} 1, & Y_n \geq q, \\ 0, & Y_n < q. \end{cases} \quad (8)$$

It is a Bernoulli random variable that has the probability distribution characterized by the random variable Y_n , which is drawn from the distribution described in Eq. (7). The goal of reconstruction is to recover the object light response $\mathbf{v}$ from the binary measurements B_n , which is linked by the discretized light flux φ_n at the hologram plane. Thus, the probability distribution of B_n is first expressed as

$$\Pr[B_n = 1] = \int_q^{\infty} p_{Y_n}(y) dy \quad (9)$$

$$= \int_q^{\infty} \sum_{k=0}^{\infty} \frac{\varphi_n^k e^{-\varphi_n}}{k!} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-k)^2}{2\sigma^2}} dy \quad (10)$$

$$= \sum_{k=0}^{\infty} \frac{\varphi_n^k e^{-\varphi_n}}{k!} \Phi\left(\frac{k-q}{\sigma}\right), \quad (11)$$

where $\Phi(\cdot)$ is the cumulative density function of the standard normal distribution.

In this study, we investigate the system with a uniform quantization threshold $q = 1$. While the optimal threshold design is an emerged topic, which could potentially improve the performance of QIS devices, especially in high dynamic range imaging,³¹ we choose to follow a standard configuration for the sake of consistency and comparability with the existing literature.^{18–20} While this research direction is not included in this study, it is certainly worthy of further investigation. Previous computational methods of QIS signal processing focus on the Poisson component by ignoring the read noise.^{28,31} We follow the same assumption in this study. However, it is worth noting that we test our algorithms on the realistic model considering the influence of different read noise levels. In our implementation (during training), we assume that $\sigma = 0$. This gives us

$$\Pr[B_n = 1] = \sum_{k=1}^{\infty} \frac{\varphi_n^k e^{-\varphi_n}}{k!} = 1 - e^{-\varphi_n}, \quad (12)$$

$$\Pr[B_n = 0] = 1 - \Pr[B_n = 1] = e^{-\varphi_n}. \quad (13)$$

Then, given the discretized object light response for the 3D light field $\mathbf{v} \in \mathbb{C}^Q$, following the mapping between $\mathbf{v}$ and φ_n in Eq. (5), the probability distribution at each sensing unit is further expressed as

$$\Pr[B_n = 1] = 1 - e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}}, \quad \Pr[B_n = 0] = e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}}, \quad (14)$$

for $mK \leq n < (m+1)K$

where K is the overall sampling ratio as stated previously. Therefore, due to the identical and independent measurements at each sensing unit, a QIS outputs a signal $\mathbf{b} = [b_1, b_2, \dots, b_N]$ with the joint distribution as

$$f(\mathbf{b}|\mathbf{v}) = \prod_{n=1}^N \Pr[B_n = b_n|\mathbf{v}] = \prod_{m=1}^M \prod_{k=1}^K \left[(1 - b_{mK+k}) e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}} + b_{mK+k} \left(1 - e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}} \right) \right]. \quad (15)$$

Since b_n is a binary value, introducing two functions

$$K_m^1 = \sum_{k=1}^K b_{mK+k}, \quad K_m^0 = K - \sum_{k=1}^K b_{mK+k}, \quad (16)$$

we can write

$$f(\mathbf{b}|\mathbf{v}) = \prod_{m=1}^M \left(e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}} \right)^{K_m^0} \left(1 - e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}} \right)^{K_m^1}, \quad (17)$$

which is the forward model for the entire imaging system.

B. Physics-informed network

To reconstruct the object wavefront information from the QIS recordings, the inverse problem approach is used, which aims to find the optimal object that would produce the observed output

while satisfying some physical constraints. Concretely, given the forward measurement model $f(\mathbf{b}|\mathbf{v})$ in Eq. (17), the optimal object is estimated by the *maximum a posteriori* (MAP) solution of a Bayesian framework with the prior constraint $p(\mathbf{v})$,

$$\mathbf{v}^* = \arg \max_{\mathbf{v}} f(\mathbf{b}|\mathbf{v}) + \mu p(\mathbf{v}), \quad (18)$$

where μ is used to control the relative effect of the regularizer on the reconstruction. Substituting Eq. (17) into Eq. (18) and taking the negative logarithm operation to transform the product term into summation, the MAP estimate solution of the proposed imaging modality is derived as

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} \sum_{m=1}^M \left[K_m^0 \frac{[\mathbf{H}\mathbf{v}]_m}{K} - K_m^1 \log \left(1 - e^{-\frac{[\mathbf{H}\mathbf{v}]_m}{K}} \right) \right] - \mu \log p(\mathbf{v}). \quad (19)$$

Inspired by the previous work for Poisson denoising,²⁵ the forward and prior terms are decoupled by introducing two independent auxiliary variables $\boldsymbol{\varphi}$ and $\boldsymbol{\alpha}$, which reformulate the original minimization problem into the constrained format,

$$\mathbf{v}^* = \arg \min_{\mathbf{v}} F(\boldsymbol{\varphi}) + \mu s(\boldsymbol{\alpha}) \quad \text{s.t.} \quad \boldsymbol{\varphi} = \mathbf{H}\mathbf{v}, \quad \boldsymbol{\alpha} = \mathbf{v}. \quad (20)$$

For notation simplicity, we define $F(\boldsymbol{\varphi}) = \sum_{m=1}^M \left[K_m^0 \frac{\varphi_m}{K} - K_m^1 \log (1 - e^{-\frac{\varphi_m}{K}}) \right]$ and use $s(\boldsymbol{\alpha})$ to represent the task-oriented prior in

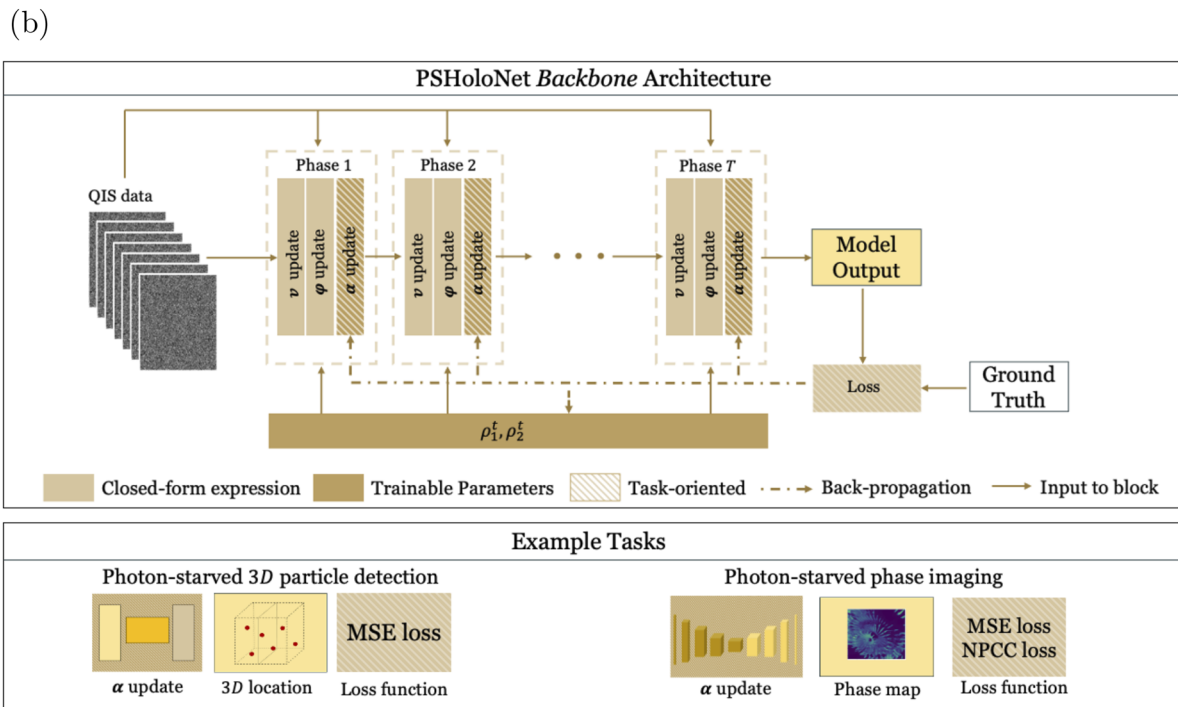
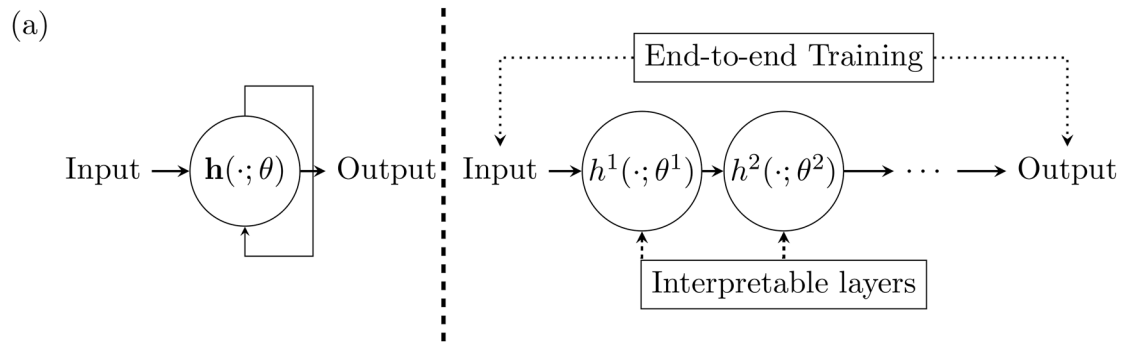


FIG. 3. Network design of PSHoloNet. (a) Conceptual schematic of unrolling an iterative algorithm (left) with handcrafted tuned parameters into its corresponding unrolled network (right) with trainable parameters. (b) PSHoloNet (top) takes the snapshot QIS recordings as input, while the output is determined by different inference applications. The task-oriented operation design is indicated by the blocks filled with the strip pattern, which are demonstrated on two common DH inference tasks (bottom), i.e., 3D particle volumetric reconstruction and phase imaging.

an implicit form, which will be introduced later. The augmented Lagrangian function for this problem is given by

$$\{\mathbf{v}^*, \boldsymbol{\varphi}^*, \boldsymbol{\alpha}^*, \mathbf{u}_1^*, \mathbf{u}_2^*\} = \arg \min_{\mathbf{v}, \boldsymbol{\varphi}, \boldsymbol{\alpha}, \mathbf{u}_1, \mathbf{u}_2} \left[F(\boldsymbol{\varphi}) + \mu s(\boldsymbol{\alpha}) + \frac{\rho_1}{2} \|\boldsymbol{\varphi} - \mathbf{H}\mathbf{v} + \mathbf{u}_1\|^2 - \frac{\rho_1}{2} \|\mathbf{u}_1\|^2 + \frac{\rho_2}{2} \|\boldsymbol{\alpha} - \mathbf{v} + \mathbf{u}_2\|^2 - \frac{\rho_2}{2} \|\mathbf{u}_2\|^2 \right], \quad (21)$$

where $\mathbf{u}_1, \mathbf{u}_2$ are the scaled Lagrangian multipliers and ρ_1, ρ_2 are the penalty parameters associated with two constraints in Eq. (20). The minimizer of Eq. (20) is the saddle point of Eq. (21),³² which is found by repeatedly performing the following alternating direction method of multipliers (ADMM)³² steps until convergence:

$$\mathbf{v}^{t+1} = \arg \min_{\mathbf{v}} \frac{\rho_1}{2} \|\boldsymbol{\varphi}^t + \mathbf{u}_1^t - \mathbf{H}\mathbf{v}\|^2 + \frac{\rho_2}{2} \|\boldsymbol{\alpha}^t + \mathbf{u}_2^t - \mathbf{v}\|^2, \quad (22)$$

$$\boldsymbol{\varphi}^{t+1} = \arg \min_{\boldsymbol{\varphi}} F(\boldsymbol{\varphi}) + \frac{\rho_1}{2} \|\boldsymbol{\varphi} - (\mathbf{H}\mathbf{v}^{t+1} - \mathbf{u}_1^t)\|^2, \quad (23)$$

$$\boldsymbol{\alpha}^{t+1} = \arg \min_{\boldsymbol{\alpha}} \mu s(\boldsymbol{\alpha}) + \frac{\rho_2}{2} \|\boldsymbol{\alpha} - (\mathbf{v}^{t+1} - \mathbf{u}_2^t)\|^2, \quad (24)$$

$$\mathbf{u}_1^{t+1} = \boldsymbol{\varphi}^{t+1} - \mathbf{H}\mathbf{v}^{t+1} + \mathbf{u}_1^t, \quad (25)$$

$$\mathbf{u}_2^{t+1} = \boldsymbol{\alpha}^{t+1} - \mathbf{v}^{t+1} + \mathbf{u}_2^t. \quad (26)$$

However, the requirement of hundreds to thousands of iterations to converge makes it time-consuming and not suitable for real-time imaging deployment. Moreover, the sensitivity of hand-tuned parameters, calibration errors, and reconstruction artifacts resulting from model mismatch degrades its performance in real-world tasks. Therefore, the one-time training scheme with automatic parameters updating is preferred, as featured by the data-driven methods. However, such methods suffer from less interpretability, which is essential in identifying the potential failures of an imaging system. Thus, to integrate the merits of the two approaches, we adapt the algorithm unrolling strategy to combine the model-based and data-driven methods.^{33–35}

The conceptual illustration of such a method is shown in Fig. 3(a). The iterative process (left) is unfolded into an end-to-end trainable network (right). Specifically, each updating step in the iteration is transferred into one layer that is cascaded in the deep network. Passing through the network is equivalent to executing the iterative algorithm a finite number of times. In this way, the algorithm parameters $\boldsymbol{\theta}$ (such as the model parameters and regularization coefficients) are transferred to the trainable network parameters $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \dots$, which are automatically tuned via back-propagation under supervised learning.

Following this scheme, we design an end-to-end trainable network for the photon-starved holographic imaging system (PSHoloNet), which is the unrolled version of the iterative process in Eqs. (22)–(26). Its backbone structure is shown in Fig. 3(b), which consists of several blocks. Each block contains three layers corresponding to the subproblems in Eqs. (22)–(24), respectively. The network input is the snapshot QIS recordings, while the output differs for different inference applications.

As shown in Fig. 3(b), we use the closed-form expression for $\mathbf{v}$ -update that is defined in Eq. (22). It is a least-square estimate problem, which has a closed-form solution. In order to avoid matrix inversions, the fast Fourier transform is used for an efficient implementation as

$$\mathbf{v}^{t+1} = \mathcal{F}^{-1} \left[\frac{\mathcal{F}(\tilde{\mathbf{v}}_1^t) + (\rho_2/\rho_1) \overline{\mathcal{F}(\mathbf{h})} \mathcal{F}(\tilde{\mathbf{v}}_2^t)}{1 + (\rho_2/\rho_1) |\mathcal{F}(\mathbf{h})|^2} \right], \quad (27)$$

where $\mathcal{F}/\mathcal{F}^{-1}$ represents the forward/inverse discrete Fourier transform and $\mathbf{h} = [h_{z_1}, h_{z_2}, \dots, h_{z_L}]$ is the concatenated propagation kernels at different depths defined in Eq. (2). For notation simplicity, we define $\tilde{\mathbf{v}}_1^t = \boldsymbol{\varphi}^t + \mathbf{u}_1^t$ and $\tilde{\mathbf{v}}_2^t = \boldsymbol{\alpha}^t + \mathbf{u}_2^t$.

For $\boldsymbol{\varphi}$ -update described in Eq. (23), it is a proximal mapping of $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \dots, \varphi_N]$ associated with the negative likelihood for a truncated Poisson distribution. Defining $\tilde{\boldsymbol{\varphi}}^t = \mathbf{H}\mathbf{v}^{t+1} - \mathbf{u}_1^t$ transforms Eq. (23) into

$$\begin{aligned} \boldsymbol{\varphi}^{t+1} &= \arg \min_{\boldsymbol{\varphi}} F(\boldsymbol{\varphi}) + \frac{\rho_1}{2} \|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}^t\|^2 \\ &= \arg \min_{\boldsymbol{\varphi}} \sum_{m=1}^M \left[K_m^0 \frac{\varphi_m}{K} - K_m^1 \log \left(1 - e^{-\frac{\varphi_m}{K}} \right) \right] \\ &\quad + \frac{\rho_1}{2} \|\boldsymbol{\varphi} - \tilde{\boldsymbol{\varphi}}^t\|^2. \end{aligned} \quad (28)$$

It is a separable convex optimization problem that could be solved without iteration.³⁶ For the m -th pixel, the optimal is acquired by finding the root of the equation where the gradient is set to zero,

$$K e^{-\frac{\varphi_m}{K}} (1 + \rho_1 (\varphi_m - \tilde{\varphi}_m^t)) = K_m^0 + \rho_1 K (\varphi_m - \tilde{\varphi}_m^t). \quad (29)$$

The $\boldsymbol{\alpha}$ -update layer at the t -th block is the solution of Eq. (24), which is the proximal point of $\tilde{\boldsymbol{\alpha}}^t = \mathbf{v}^{t+1} - \mathbf{u}_2^t$ with respect to the prior term $s(\boldsymbol{\alpha})$. Therefore, it is customized to make PSHoloNet adaptive for various inference tasks.

For 3D particle volumetric reconstruction, the network output is the 3D location of the particles. The particle field is generally assumed to be real-valued with physically realistic sparsity properties in a linear basis.^{6,37} Therefore, the $\boldsymbol{\alpha}$ -update is designed to introduce sparsity regularization. We consider a transform $\mathcal{T}(\cdot)$ that sparsifies the layer input $\tilde{\boldsymbol{\alpha}}^t$. To avoid handcrafted parameters for constructing such a sparsity basis, the convolutional neural network (CNN) blocks are used with their generalized approximation property.³⁸ The sparsity of the transformed signal $\mathcal{T}(\tilde{\boldsymbol{\alpha}}^t)$ is encouraged by the soft-threshold operation on the transformed domain. The inverse of $\mathcal{T}(\cdot)$, denoted by $\mathcal{T}^{-1}(\cdot)$, is then introduced to recover the signal back to its original signal domain. Moreover, it takes the same structure with $\mathcal{T}(\cdot)$ for a consideration of structure symmetry. Such a symmetric constraint is further emphasized by incorporating it into the loss function during the network training (see Sec. II C). Therefore, $\boldsymbol{\alpha}^t$ can be efficiently computed as

$$\boldsymbol{\alpha}^t = \mathcal{T}^{-1}(\text{soft}(\mathcal{T}(\tilde{\boldsymbol{\alpha}}^t), \theta^t)), \quad (30)$$

where θ^t is the shrinkage threshold and is a trainable parameter at the t -th block. The detailed structures are presented in the Appendix.

For phase imaging, the output α^t from the α -update layer at the t -th block can be interpreted as a denoising process of the layer input $\tilde{\alpha}^t$. Such an image-to-image registration can be rephrased as

$$\alpha^t = \mathcal{D}(\tilde{\alpha}^t). \quad (31)$$

Considering its hourglass structure and prominent performance in various image reconstruction problems, we adapt the widely used U-Net structure with the residual connection.³⁹ It is composed of a set of downsampling and upsampling operations, of which each has a skip connection between the resulting feature blocks of the same size. This reduces the semantic gap between the encoder and decoder features. The residual blocks are included within each operator, which aims to boost the network's ability to overcome vanishing gradients. The detailed structures are presented in the Appendix. To increase the network flexibility, the penalty parameters ρ_1 and ρ_2 are varied across different phases that are determined in one-time training.

C. Network implementation

For different interference tasks, the loss function is designed to consider the data fidelity and prior constraints simultaneously. The data fidelity penalty for each block is considered by minimizing the mean-square-error (MSE) loss function,

$$\mathcal{L}_d^t = \|\alpha^t - \mathbf{v}^{gt}\|_2^2, \quad (32)$$

where α^t stands for the output of the t -th phase and $\mathbf{v}^{gt}$ stands for the ground truth.

As for the prior constraint, in 3D particle volumetric reconstruction, the symmetry between the transform $\mathcal{T}(\cdot)$ and its inverse $\mathcal{T}^{-1}(\cdot)$ in the sparsity-inducing prior is enforced by

$$\mathcal{L}_p^t = \|\mathcal{T}^{-1}(\mathcal{T}(\tilde{\alpha}^t)) - \tilde{\alpha}^t\|_2^2. \quad (33)$$

In phase imaging, since it is rephrased as an image-to-image reconstruction, the correlative metric, such as the negative Pearson correlation coefficient (NPCC),⁴⁰ is included for preserving the global similarity between the output and its target equivalents, which is defined as

$$\mathcal{L}_p^t = -\frac{\text{cov}(\alpha^t, \mathbf{v}^{gt})}{\sigma_{\alpha^t} \sigma_{\mathbf{v}^{gt}}}, \quad (34)$$

where $\text{cov}(\alpha^t, \mathbf{v}^{gt})$ is the covariance between the prediction α^t at the t -th stage and the ground truth phase map $\mathbf{v}^{gt}$, and σ_{α^t} and $\sigma_{\mathbf{v}^{gt}}$ are their standard deviation, respectively.

Therefore, the total task-oriented loss function is the weighted sum of the data-fidelity penalty and task-oriented prior constraint,

$$\mathcal{L}_{total} = \frac{1}{T} \sum_{t=1}^T (\mathcal{L}_d^t + \gamma \mathcal{L}_p^t), \quad (35)$$

where T stands for the total number of phases in the unrolled network and γ is the regularization parameter. In our experiments, γ is empirically set as 0.01 and $T = 5$ considering the convergence speed and computational cost. In both applications, the network is implemented using Python, version 3.7.0, with the PyTorch framework,

version 1.11.1. We use a single NVIDIA GeForce RTX 3070 GPU to accelerate the training process. The Adam optimizer⁴¹ is used with an initial learning rate of 0.0001, which can handle the sparse gradients and realize efficient computation.

III. RESULTS AND DISCUSSION

We demonstrate photon-starved snapshot holography using a simple Gabor holographic setup equipped with a QIS as the recording device. To enable 3D particle volumetric reconstruction and phase imaging, we implement two variants of a modular network, which we term PSHoloNet, as described in Sec. II B. We first mathematically model the imaging system and investigate its performance via simulation. Then, we verify the system using experimental holograms of particle volumes and analyze the influence of read noise, which is unavoidable in real sensors. The results are presented in Secs. III A–III C.

A. Reconstruction of synthetic 3D particle volume

A photon-starved condition with less than one photon per pixel is created by setting the mean rate of Poisson statistics of photon arrival to vary between 0 and 1. The QIS, which consists of an array of photodetectors called jots, detects photon arrivals following the “uniform-binning” scheme. This device is considered an oversampling mechanism with an overall ratio K , which is comprised of spatial oversampling K_s and temporal oversampling K_T (refer to Sec. II A for more details). In this study, we set $K_s = 4$ and $K_T = 100$, in accordance with the established practice.^{24,31} This configuration is equivalent to an integration time of 0.1 s under a 1 kHz frame rate.²¹ By utilizing this setup, we are able to obtain accurate and reliable measurements of photon arrivals with a larger field-of-view that supports a snapshot capture of the whole hologram without the mechanical scanning process.

We now consider a volume of 3D particles as the target object. The particles are assumed to be pure amplitude disks with a diameter of 20 μm , which are randomly spread in a volume with a distance of 5–7 cm from the hologram plane. It is discretized to $256 \times 256 \times 7$ voxels, with an xy resolution of 10 μm . Following the image formation model of the QIS-DH system, the holograms generated on the sensor plane are normalized and set as the mean rate of the Poisson arrival, in accordance with the low-flux photon condition (<1 ppp). We incorporate this physical model into an end-to-end training framework to construct the PSHoloNet, in order to detect the volumetric distribution of the 3D particles. We incorporate this physical model into an end-to-end training framework to construct the PSHoloNet, in order to detect the volumetric distribution of the 3D particles. The physical model provides *a priori* knowledge of the underlying system, which can be used to constrain the network and further improve its learning effectiveness that supports a small training dataset. Considering the complexity of the network and the size and diversity of the datasets, we generate 500 holograms and we follow the same typical data-splitting arrangement with a ratio of 7:2:1 for training, validation, and testing datasets. We train the network for 300 epochs and cost 4 h in total, where the loss function converges to a stable value and further training does not improve the performance of the network significantly. We find that it is

sufficient to effectively train the network and achieve high accuracy in our tests. After the network training phase, which only needs to be done once, the trained PSHoloNet is capable of processing a single snapshot QIS recording in just 0.04 s.

We conduct a comparison of our system with a state-of-the-art method called OSNet,⁷ which does not take into account the physical model. The visual results are given in Fig. 4, where the green circles are the ground truth and the red triangles represent the prediction results from the tested algorithms. As expected, under sufficient illumination where the interference patterns are recorded with high visibility [Fig. 4(b)], OSNet succeeds in restoring the 3D particles distribution in Fig. 4(e). However, under low light [Fig. 4(c)], the captured hologram is heavily corrupted by the intrinsic read noise in the sensor (i.e., $\sigma = 2e^-$), which results in the barely recognizable patterns and, in turn, a poor performance of OSNet [Fig. 4(f)]. In contrast, the binary recording nature of QIS leads to a negligible amount of read noise, and the dominant factor affecting fringes quality is the randomness of the photon arrivals [Fig. 4(d)]. By carefully taking this into account, PSHoloNet successfully reconstructs the particle volume as shown in Fig. 4(g). Numerically, it has a small position error of only 1.26 voxels, compared to the very significant errors shown in Fig. 4(f).

In order to further investigate the robustness of our system against particle density, the trained network is tested under various particle concentrations ranging from 1×10^{-4} to 6×10^{-4} particles per volume (ppv). Since the target distribution (ground truth) is available in the simulation, the quantitative evaluation can

be made by considering two common metrics: (1) recall, which is calculated as

$$R = \frac{N_t}{N_t + N_m}, \quad (36)$$

where N_t is the number of correct predictions and N_m is the number of missed ground truths; (2) precision, which is calculated as

$$P = \frac{N_t}{N_t + N_f}, \quad (37)$$

where N_f is the number of wrong inferences.

The reconstruction results presented in Fig. 5 and Table I provide compelling evidence for the robustness of PSHoloNet in the face of particle density variation. In the 3D plot shown in Fig. 5, because the number of wrong and missed predictions is relatively low compared to the number of correct predictions, they may be obscured by the large number of correct detections in the plot and hardly discernible. Therefore, we use the precision and recall metrics in Table I to take into account the number of both correct and incorrect predictions quantitatively made by our system. As indicated, our system is able to accurately identify particles with a high degree of confidence. In addition, despite a sixfold increase in the number of particles from 50 to 300, PSHoloNet continues to deliver accurate predictions of particle locations. Across different density levels, PSHoloNet achieves an average recall of 93.6% and a preci-

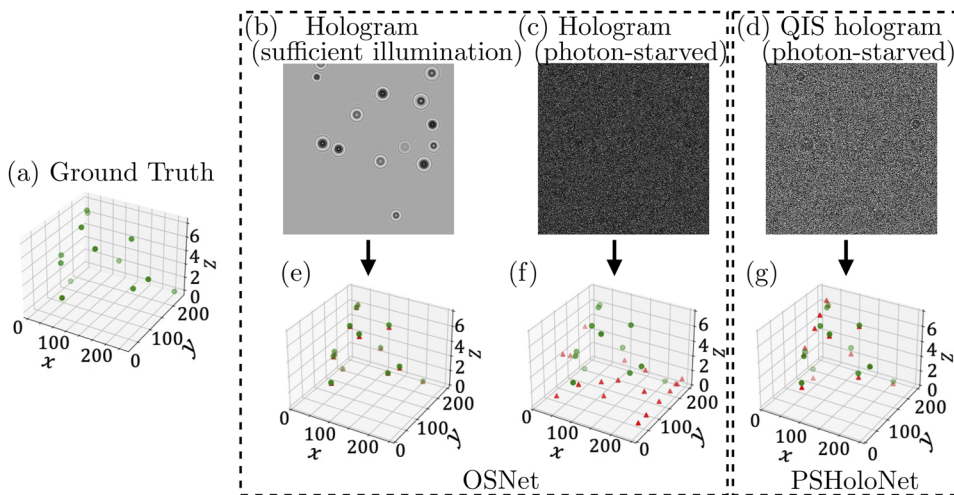


FIG. 4. Reconstruction result comparison of 3D particle volume using different imaging systems. Green circles: The ground truth of particle distribution. Red triangles: The prediction results from the tested algorithms. (a) Ground truth of the particle distribution. The reconstruction results from OSNet⁷ with holograms under (b) sufficient illumination and (c) photon-starved condition are shown in (e) and (f), respectively; the reconstruction result from PSHoloNet with (d) QIS-recorded hologram under photon-starved condition is shown in (g).

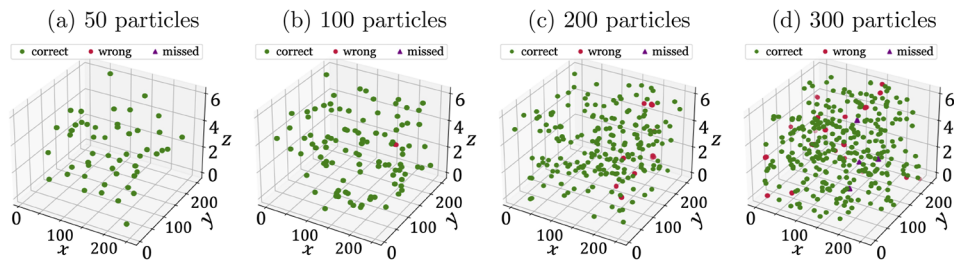


FIG. 5. Reconstruction of 3D particle volume at different densities. The green, red, and purple symbols represent the correct, wrong, and missed predictions, respectively.

TABLE I. Evaluation results for particle extraction metrics and prediction errors at different particle densities.

Particle density (ppv)	Recall (%)	Precision (%)	XY error (μm)	Z error (mm)
1×10^{-4}	99.6	98.5	23.3	0.031
2×10^{-4}	95.3	98.2	23.4	0.039
4×10^{-4}	87.2	99.5	23.6	0.085
6×10^{-4}	92.3	90.7	24.4	0.137

sion of 96.73%. In order to make a quantitative assessment of the particle location prediction, we follow the common practice in object detection tasks and calculate the deviation. First, a buffer area is pre-defined around each ground-truth particle. Then, the predicted particle from the network that falls within this area is considered a matched particle, and the difference between their coordinates is taken as the prediction error in spatial distance. In our study, the buffer area is defined as a sphere with a five-voxel radius, and the results are shown in Table I. Specifically, for the 300-particle case, which is the densest situation, the maximum prediction error in correctly matched particle pairs is $24.4 \mu\text{m}$ along the xy-plane and 0.137 mm along the z axis. These values account for only 0.697% and 0.685% of the total range, respectively.

B. Phase imaging of synthetic objects

We investigate the application of the proposed network in photon-starved phase imaging by replacing the object field with pure phase objects. They are simulated by mapping the intensity patterns

from some public natural images⁴² to a phase modulation ranging from $-\pi$ to π , as shown in Fig. 6(a). The transmissive phase object is considered at a distance of 2 cm in front of the sensor plane with the same configuration for the illumination source and QIS. We simulate 300 holograms and follow the same training strategy as Sec. III A.

Figures 6(b) and 6(c) show examples of phase imaging under photon-starved situations in a conventional DH system. Again, the randomness of photon arrivals and intrinsic read-noise result in highly contaminated interference patterns. Under this situation, even a state-of-the-art iterative reconstruction algorithm called deep compressed object decoder (DCOD)⁴³ fails to recover the phase information. On the other hand, considerable phase information is successfully captured in the QIS-recorded holograms, and PSHoloNet is able to recover the phase images.

Such adaptability of this system is achieved by designing and incorporating task-oriented priors and loss functions. In previous 3D particle volumetric reconstruction tasks, since the particle field is inherently real-valued due to the relatively small particle size

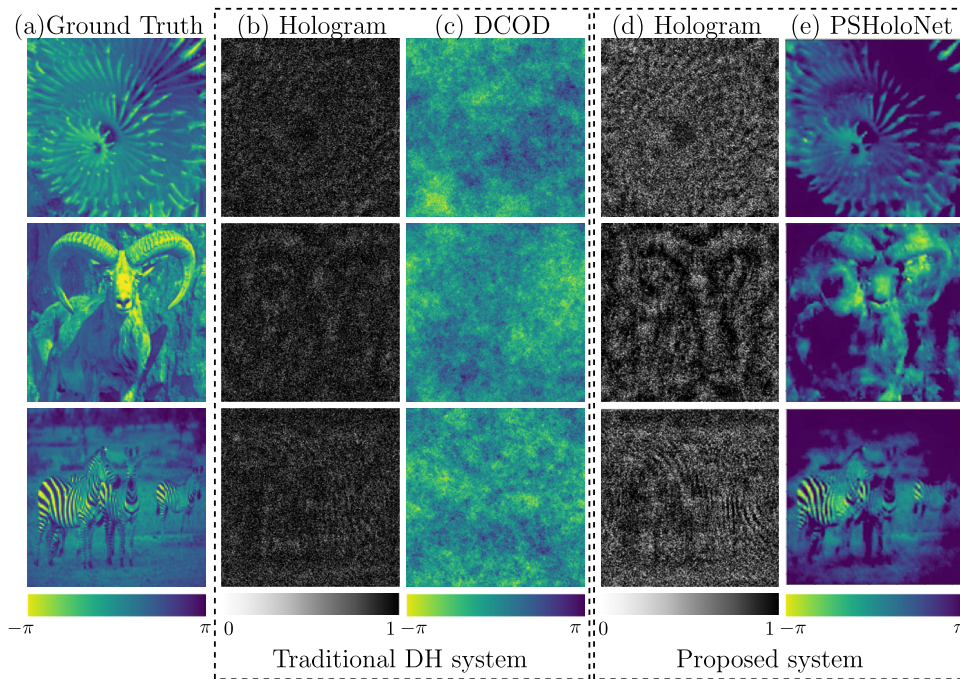


FIG. 6. Phase imaging comparison between the traditional DH system and the proposed system under the photon-starved situation. (a) The ground-truth phase images from Ref. 42. (b) The holograms recorded by the traditional sensor. We assume that the photon level is $<1 \text{ ppp}$. The sensor has a read noise of $\sigma = 2e^-$ and operates at 8 bits. Shown in this figure is one image frame. (c) The reconstruction results from DCOD.⁴³ (d) The holograms recorded by QIS. We use $K = 400$ frames, where each frame is binary and the read noise is $\sigma = 0.19e^-$ following Ref. 44. (e) The reconstruction result from PSHoloNet.

(micro-/nano-scale), the use of mean-square-error (MSE) loss is straightforward with fast convergence and efficiency. However, for phase imaging, since there are much richer spatial (or spectral) details of the object, the selection of loss function has a significant influence on the reconstruction performance. To investigate such an effect, we further train PSHoloNet separately with the same synthetic phase object dataset while using different loss functions for the back-propagation and parameter updating. From the visual inspection shown in Fig. 7, undesired checkerboard artifacts are observed in the output images from the network trained purely by MSE loss. It reveals the insufficiency of retrieving the phase maps only from the pixel-wise comparison. On the contrary, the negative Pearson correlation (NPCC) loss preserves the spatial correlations between the neighbor pixels, which increases the structural similarity between the network prediction and the target feature map. As shown in Fig. 7, it leads to better reconstructed phase maps with more informative and distinguishable features. To quantify this comparison, we plot the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) values between the reconstructed results and their ground-truth equivalents along the training process in Fig. 7(b). As indicated, only using MSE as the loss function results in poor structure recovery (an SSIM of 0.1). By introducing the NPCC loss, the more structural similarity is preserved in the predicted phase maps with a three times improvement (an SSIM of 0.32).

C. Experimental particle reconstruction with read noise analysis

We now proceed to verify the above experimentally. To prepare the particle volume sample for our experiment, we disperse 50 μm polystyrene particles in a physiological saline solution, which covers a distance ranging from 10 to 60 mm from the sensor plane. This involves thoroughly mixing the plastic particles with the saline solution until the particles are evenly distributed throughout the solution. Due to the similar density of the particles and the liquid, the particles diffuse in the solution and exhibit Brownian motion. This sample preparation method allows us to study the behavior and trajectories of the plastic particles from consecutive holograms. The illumination source is a 660 nm laser that is collimated to create the planar wave. The photon level is controlled by inserting a neutral density filter between the light source and the sample. A Sony IMX264 CMOS sensor with a pixel pitch of 3.45 μm is used to record the holograms, and computationally converted to QIS data, following the same mathematical modeling in Sec. III A that has been experimentally verified extensively in the literature.^{20,26,30}

Since there is no ground truth, the inference using the state-of-the-art OSNet hologram reconstruction method with the ideal imaging condition is used as the reference for performance evaluation. The prediction result example is shown in Fig. 8, with an

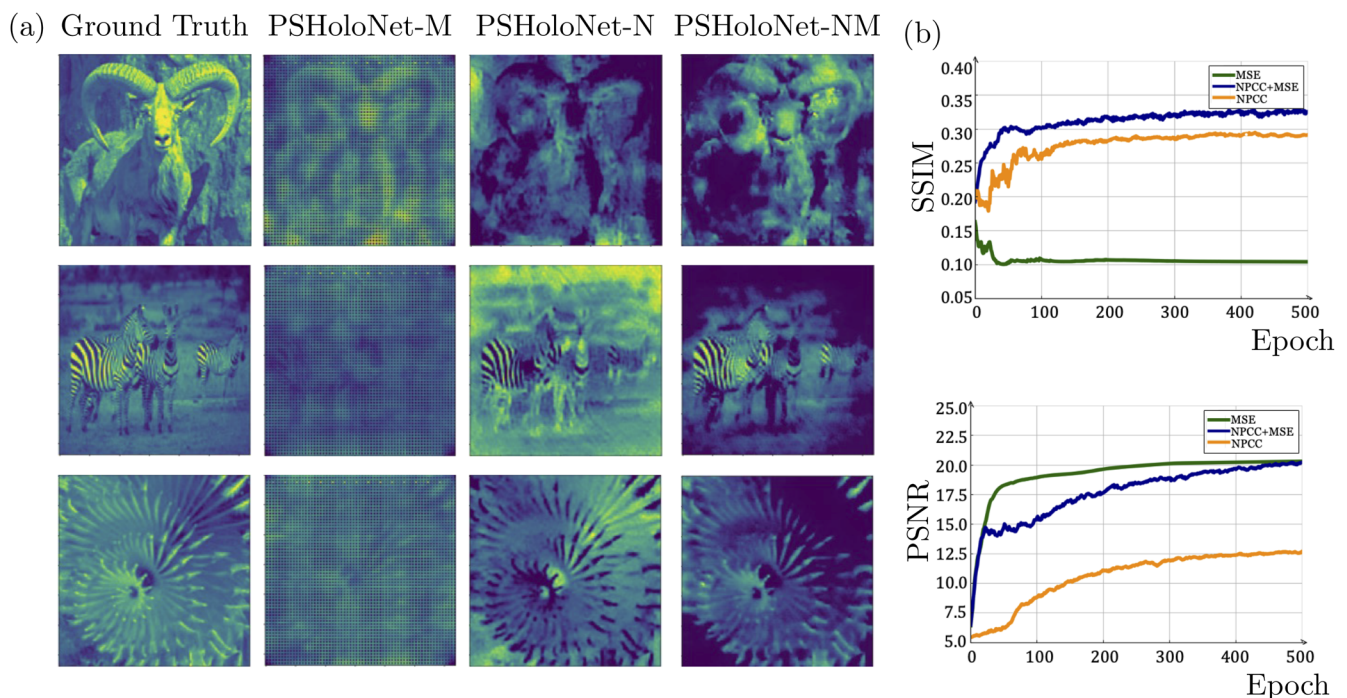


FIG. 7. Impact of loss functions. The condition of NPCC and MSE offers the best performance. (a) Reconstruction results from PSHoloNet trained with mean-square-error loss (PSHoloNet-M), negative Pearson correlation coefficient loss (PSHoloNet-N), and their combination (PSHoloNet-NM). (b) The SSIM and PSNR curves for each case, respectively.

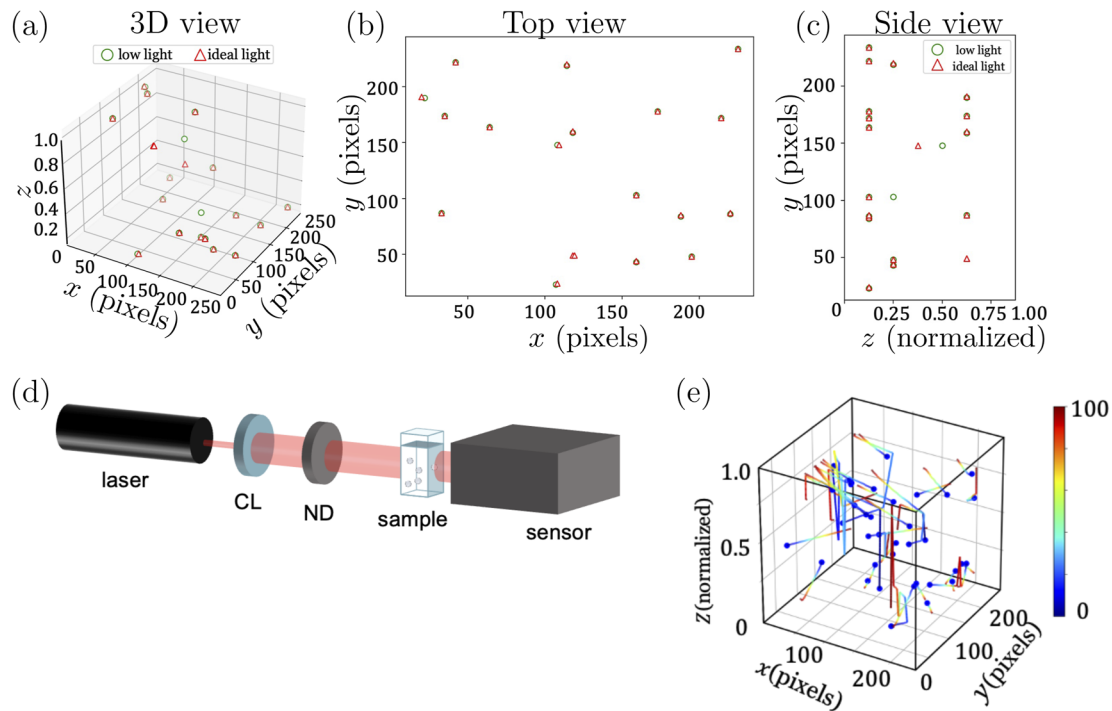


FIG. 8. Particle volumetric reconstruction on the experimental holograms. The visual inspection is provided with (a) a 3D view, (b) a top view, and (c) a side view. Green circles: The inference results under the photon-starved condition using PSHoloNet. Red triangles: The inference results under the ideal light condition using the previous reconstruction method (OSNet).⁷ (d) The schematic illustration of the experimental setup that we used in our experiment. ND: neutral density filter. CL: collimating lens. (e) Trajectories of the selected particles. The color of the trajectories indicates the frame index.

82.6% recall and an 88.1% precision on average. The slight decrease in accuracy levels compared with the simulation is acceptable due to the inevitable misalignment errors in optical components in the real experimental setup. Based on the volumetric reconstruction on consecutive frames, the particle tracking is conducted to get the trajectories as shown in Fig. 8(e). The reported mean lateral and axial errors are $11.2 \mu\text{m}$ and 0.137 mm , respectively. The precision is competitive with that achieved by OSNet under the ideal lighting condition.

In our results reported so far, we have verified the feasibility of the proposed QIS-DH system to recover wavefront information under photon-starved scenarios. We study the possibility of snapshot DH at an extremely low signal budget ($<1 \text{ ppp}$). Under such conditions, the signal is so small that the Poisson random variable dominates, which is highly contaminated by the intrinsic read-out noise in traditional sensors. In contrast, QIS has a prominent reduction in read noise and provides more faithful access for the Poisson randomness of the photon arrival. However, it still cannot be free of it since there is no perfect sensor.

To investigate the influence of read noise on our system, we simulate QIS-recorded holograms for particle volumes under various levels of read noise. The mean photon flux is kept below 1 ppp , and the read noise variance σ increases from $0.1e^-$ to $2e^-$. The quantitative metrics (i.e., recall and precision) are shown in Fig. 9 in the way of a box-and-whisker plot. For metric at each noise level, the box

extends from the Q1 to Q3 quartile values with a line at the median (Q2) and the whiskers show their range. The observation is that regardless of the read noise level (at least for $\sigma \leq 1e^-$), our system demonstrates robust and competitive prediction performance both

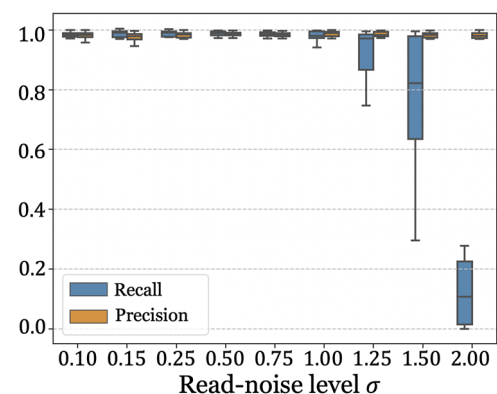


FIG. 9. Boxplot for particle volumetric reconstruction metrics under different noise levels. At each noise level, recall and precision results are depicted in the way of blue and orange box-and-whisker plots, respectively. The box extends from the Q1 to Q3 quartile values with a line at the median (Q2). The whiskers show the range of the metric.

in recall and precision (over 95%). As the noise level goes up, the number of missed particles increases, which results in a decreased recall (average of 18% under $\sigma = 2e^-$). Its dispersion also shows the unsteady performance for particle volumetric reconstruction under the photon-starved situation when the read noise is too large as is the case in the traditional sensors. On the contrary, a QIS has a relatively small value of $\sigma = 0.19e^-$, which is sufficient to guarantee its insensitivity to read noise.⁴⁴

IV. SUMMARY

In this study, we demonstrate the integration of QIS in developing a snapshot photon-starved DH system. Enabled by the physics-informed PSHoloNet, our approach enables 3D particle volumetric reconstruction and phase imaging under a limited photon budget by single-bit holograms. This facilitates more advanced holography applications in situations when the scarcity of photons is essential and unavoidable, which has numerous applications in fields such as microscopy, biomedical imaging, and materials science. In addition, we have shown that QIS can be used as an effective imaging modality in scientific and life science imaging applications. Our findings open up new possibilities for QIS in this area, and we anticipate that it will encourage new research initiatives and collaborations, leading to the development of more advanced and efficient imaging techniques that can benefit the scientific community.

ACKNOWLEDGMENTS

This work was supported by the Research Grants Council of Hong Kong (Grant Nos. GRF 17200019, GRF 17201620, and GRF 17201822).

The authors would like to thank Dr. Ni Chen (University of Arizona, USA) for her comments on an early version of this project and suggestions on the experimental setup.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yunping Zhang: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Stanley H. Chan:** Methodology (supporting); Writing – review & editing (supporting). **Edmund Y. Lam:** Conceptualization (supporting); Funding acquisition (lead); Supervision (lead); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: STRUCTURE OF α -UPDATE BLOCK

As introduced in Sec. II, α -update can be treated as a Gaussian denoiser for data with prior information, which are task-oriented and integrated by using the powerful representation of CNNs with different structures. Different structures used in this project are presented in the following.

1. α -update for the reconstruction of 3D particles distribution

As shown in Fig. 10, the sparsity regularization is performed by the soft-threshold operation on the transformed domain. The CNN blocks are used with their generalized approximation property to represent the non-linear transform $\mathcal{T}(\cdot)$.³⁸ The symmetric inverse transform $\mathcal{T}^{-1}(\cdot)$ from the sparsity domain is designed as the same structure. Specifically, each CNN block contains consecutive convolutional layers separated by a batch normalization layer and leaky rectified linear (ReLU) activation layer, which is a widely used structure with promising performance in various vision tasks.^{6,38,42} Specifically, 3×3 convolutional layers are used with the same volume depth. A skip connection is made from the input directly to the final activation layer. Overall, the update process in Eq. (31) is rephrased as

$$\alpha^t = \mathcal{T}^{-1}(\text{soft}(\mathcal{T}(\tilde{\alpha}^t), \theta^t)), \quad (\text{A1})$$

where $\tilde{\alpha}^t = \mathbf{v}^{t+1} - \mathbf{u}_2^t$ and θ^t is the shrinkage threshold, which is a trainable parameter at the t -th phase.

2. α -update for the reconstruction of the phase map

Figure 11 shows the architecture details of the task-oriented prior in the phase map reconstruction. The input of the α -update block is the revised signal $\tilde{\alpha}^t$, and the output is the phase prediction α^t .

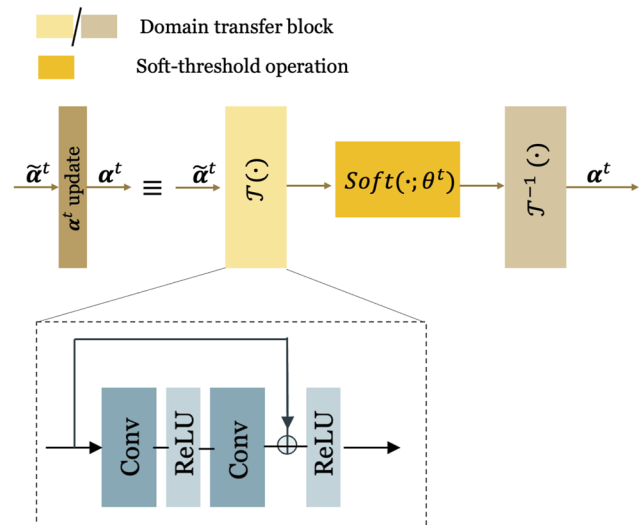


FIG. 10. α -update block for particle volumetric reconstruction task.

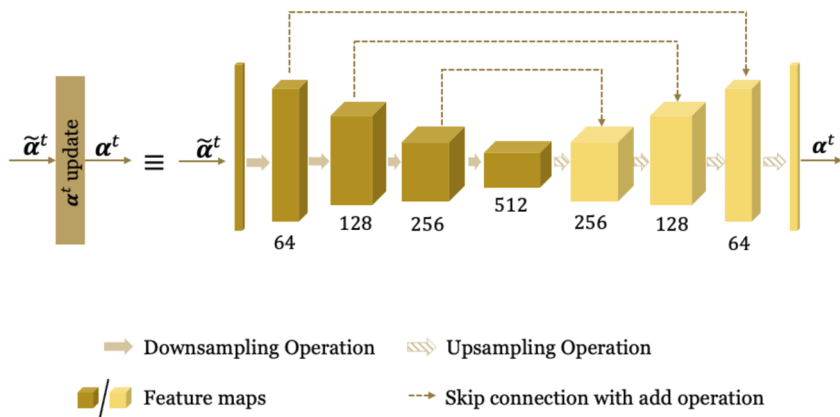


FIG. 11. α -update block for phase imaging task.

In alignment with the typical setting of U-Net, four sets of downsampling and upsampling operations are included, of which each has a skip connection between the resulting feature blocks of the same size. The number of channels for each scale is set as (64, 128, 256, 512), respectively. The residual blocks are included in the downsampling and upsampling operators, of which each is composed of a *Conv-ReLU-Conv* unit with an extra adding operation from its input to its output.

REFERENCES

- U. Schnars, C. Falldorf, J. Watson, and W. Jüptner, "Digital holography," *Digital Holography and Wavefront Sensing* (Springer, 2015), pp. 39–68.
- Y. Wu and A. Ozcan, "Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring," *Methods* **136**, 4–16 (2018).
- B. Kemper and G. Von Bally, "Digital holographic microscopy for live cell applications and technical inspection," *Appl. Opt.* **47**, A52–A61 (2008).
- V. Bianco, P. Memmolo, P. Carcagni, F. Merola, M. Paturzo, C. Distanto, and P. Ferraro, "Microplastic identification via holographic imaging and machine learning," *Adv. Intell. Syst.* **2**, 1900153 (2020).
- Y. Zhu, C. Hang Yeung, and E. Y. Lam, "Microplastic pollution monitoring with holographic classification and deep learning," *J. Phys.: Photonics* **3**, 024013 (2021).
- N. Chen, C. Wang, and W. Heidrich, "Holographic 3D particle imaging with model-based deep network," *IEEE Trans. Comput. Imaging* **7**, 288–296 (2021).
- Y. Zhang, Y. Zhu, and E. Y. Lam, "Holographic 3D particle reconstruction using a one-stage network," *Appl. Opt.* **61**, B111–B120 (2022).
- S. Wäldchen, J. Lehmann, T. Klein, S. Van De Linde, and M. Sauer, "Light-induced cell damage in live-cell super-resolution microscopy," *Sci. Rep.* **5**, 15348 (2015).
- F. Saglimbeni, S. Bianchi, A. Lepore, and R. Di Leonardo, "Three-axis digital holographic microscopy for high speed volumetric imaging," *Opt. Express* **22**, 13710–13718 (2014).
- P. A. Cheremkhin, N. N. Evtikhiev, V. V. Krasnov, V. G. Rodin, and R. S. Starikov, "Shot noise and fixed-pattern noise effects on digital hologram reconstruction," *Opt. Lasers Eng.* **139**, 106461 (2021).
- B. Mirecki, M. Rogalski, P. Arcab, P. Rogujski, L. Stanaszek, M. Józwik, and M. Trusiak, "Low-intensity illumination for lensless digital holographic microscopy with minimized sample interaction," *Biomed. Opt. Express* **13**, 5667–5682 (2022).
- K. Inoue, A. Anand, and M. Cho, "Angular spectrum matching for digital holographic microscopy under extremely low light conditions," *Opt. Lett.* **46**, 1470–1473 (2021).
- Z. Zhang, Y. Zheng, T. Xu, A. Upadhyay, Y. J. Lim, A. Mathews, L. Xie, and W. M. Lee, "Holo-UNet: Hologram-to-hologram neural network restoration for high fidelity low light quantitative phase imaging of live cells," *Biomed. Opt. Express* **11**, 5478–5487 (2020).
- P. A. Morris, R. S. Aspdin, J. E. Bell, R. W. Boyd, and M. J. Padgett, "Imaging with a small number of photons," *Nat. Commun.* **6**, 5913 (2015).
- M. Yamamoto, H. Yamamoto, and Y. Hayasaki, "Photon-counting digital holography under ultraweak illumination," *Opt. Lett.* **34**, 1081–1083 (2009).
- N. Demoli, H. Skenderović, and M. Stipčević, "Digital holography at light levels below noise using a photon-counting approach," *Opt. Lett.* **39**, 5010–5013 (2014).
- I. Kang, F. Zhang, and G. Barbastathis, "Phase extraction neural network (PhENN) with coherent modulation imaging (CMI) for phase retrieval at low photon counts," *Opt. Express* **28**, 21578–21600 (2020).
- J. Ma and E. R. Fossum, "Quanta image sensor jot with sub 0.3e- r.m.s. read noise and photon counting capability," *IEEE Electron Device Lett.* **36**, 926–928 (2015).
- Z. Yin, J. Ma, S. Masoodian, and E. R. Fossum, "Threshold uniformity improvement in 1b quanta image sensor readout circuit," *Sensors* **22**, 2578 (2022).
- E. R. Fossum, "The quanta image sensor (QIS): Concepts and challenges," in *Imaging Systems and Applications* (Optical Society of America, 2011), p. JTuE1.
- K. Morimoto, A. Ardelean, M.-L. Wu, A. C. Ulku, I. M. Antolovic, C. Bruschini, and E. Charbon, "Megapixel time-gated SPAD image sensor for 2D and 3D imaging applications," *Optica* **7**, 346–354 (2020).
- T. Al Abbas, N. Dutton, O. Almer, S. Pellegrini, Y. Henrion, and R. Henderson, "Backside illuminated SPAD image sensor with 7.83 μm pitch in 3D-stacked CMOS technology," in *2016 IEEE International Electron Devices Meeting (IEEE, 2016)*, pp. 8–1.
- P. Caramazza, A. Boccolini, D. Buschek, M. Hullin, C. F. Higham, R. Henderson, R. Murray-Smith, and D. Faccio, "Neural network identification of people hidden from view with a single-pixel, single-photon detector," *Sci. Rep.* **8**, 11945 (2018).
- A. Gnanasambandam and S. H. Chan, "HDR imaging with quanta image sensors: Theoretical limits and optimal reconstruction," *IEEE Trans. Comput. Imaging* **6**, 1571–1585 (2020).
- Y. Sanghvi, A. Gnanasambandam, and S. H. Chan, "Photon limited non-blind deblurring using algorithm unrolling," *IEEE Trans. Comput. Imaging* **8**, 851–864 (2022).
- J. Ma, S. Chan, and E. R. Fossum, "Review of quanta image sensors for ultralow-light imaging," *IEEE Trans. Electron Devices* **69**, 2824–2839 (2022).
- J. W. Goodman, *Introduction to Fourier Optics*, 4th ed. (Roberts and Company Publishers, 2017).
- F. Yang, Y. M. Lu, L. Sbaiz, and M. Vetterli, "Bits from photons: Oversampled image acquisition using binary Poisson statistics," *IEEE Trans. Image Process.* **21**, 1421–1436 (2011).

- ²⁹L. Sbaiz, F. Yang, E. Charbon, S. Susstrunk, and M. Vetterli, "The gigavision camera," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2009), pp. 1093–1096.
- ³⁰E. R. Fossum, J. Ma, S. Masoodian, L. Anzagira, and R. Zizza, "The quanta image sensor: Every photon counts," *Sensors* **16**, 1260 (2016).
- ³¹O. A. Elgendy and S. H. Chan, "Optimal threshold design for quanta image sensor," *IEEE Trans. Comput. Imaging* **4**, 99–111 (2017).
- ³²S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning* **3**, 1–122 (2011).
- ³³V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.* **38**, 18–44 (2021).
- ³⁴K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express* **27**, 28075–28090 (2019).
- ³⁵J. Sun, H. Li, Z. Xu *et al.*, "Deep ADMM-Net for compressive sensing MRI," *Advances in Neural Information Processing Systems* **29**, 1–9 (2016).
- ³⁶S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Trans. Comput. Imaging* **3**, 84–98 (2016).
- ³⁷L. Denis, D. Lorenz, E. Thiébaud, C. Fournier, and D. Trede, "Inline hologram reconstruction with sparsity constraints," *Opt. Lett.* **34**, 3475–3477 (2009).
- ³⁸J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 1828–1837.
- ³⁹K. Zhang, L. V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, 2020), pp. 3217–3226.
- ⁴⁰S. Li, M. Deng, J. Lee, A. Sinha, and G. Barbastathis, "Imaging through glass diffusers using densely connected convolutional networks," *Optica* **5**, 803–813 (2018).
- ⁴¹D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.
- ⁴²J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2009), pp. 248–255.
- ⁴³F. Niknam, H. Qazvini, and H. Latifi, "Holographic optical field recovery using a regularized untrained deep decoder network," *Sci. Rep.* **11**(10903), 1–13 (2021).
- ⁴⁴J. Ma, D. Zhang, O. A. Elgendy, and S. Masoodian, "A 0.19e⁻ rms read noise 16.7 Mpixel stacked quanta image sensor with 1.1 μ m-pitch backside illuminated pixels," *IEEE Electron Device Lett.* **42**, 891–894 (2021).