

Single-shot inline holography using a physics-aware diffusion model

YUNPING ZHANG,¹  XIHUI LIU,^{1,2} AND EDMUND Y. LAM^{1,*} 

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

²Institute of Data Science, The University of Hong Kong, Pokfulam, Hong Kong SAR, China

*elam@eee.hku.hk

Abstract: Among holographic imaging configurations, inline holography excels in its compact design and portability, making it the preferred choice for on-site or field applications with unique imaging requirements. However, effectively holographic reconstruction from a single-shot measurement remains a challenge. While several approaches have been proposed, our novel unsupervised algorithm, the physics-aware diffusion model for digital holographic reconstruction (PadDH), offers distinct advantages. By seamlessly integrating physical information with a pre-trained diffusion model, PadDH overcomes the need for a holographic training dataset and significantly reduces the number of parameters involved. Through comprehensive experiments using both synthetic and experimental data, we validate the capabilities of PadDH in reducing twin-image contamination and generating high-quality reconstructions. Our work represents significant advancements in unsupervised holographic imaging by harnessing the full potential of the pre-trained diffusion prior.

© 2024 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Holographic imaging enables wavefront reconstruction from recorded interference patterns [1], revolutionizing biomedical, physical sciences, and engineering applications [2–4]. Compared to other holographic imaging system configurations, the inline digital holography stands out for its compactness and portability, making it ideal for unique imaging requirements in on-site or field applications [5–9]. It records a hologram denoted as $y(a, b)$ at position (a, b) on the sensor plane through interference between the unscattered plane reference wave $u_r(a, b)$ and the object-scattered wave $u_o(a, b)$. Mathematically, this is expressed as

$$\begin{aligned} y(a, b) &= |u_o(a, b) + u_r(a, b)|^2, \\ &= |u_o(a, b)|^2 + |u_r(a, b)|^2 + u_r^*(a, b)u_o(a, b) + u_r(a, b)u_o^*(a, b), \end{aligned} \quad (1)$$

where $(\cdot)^*$ denotes the complex conjugation. To simplify the notation and streamline the equations, we adopt a dropout of the lateral coordinates (a, b) by representing the 2D matrix as bold symbols throughout the manuscript. In the conventional reconstruction using back-propagation (BP), due to the absence of phase information in holograms obtained through intensity-only measurements, the interchangeability of the two interfering terms u_o and u_o^* creates the undesirable twin-image artifact. It appears as an out-of-focus conjugate at the virtual image plane and degrades the quality of the reconstruction. Some attempts have been made by iteratively propagating between diverse hologram planes to retrieve the missing phase [10–13]. However, these approaches require an efficient collection of several holograms under distinct set-ups, effectively negating the primary advantages of inline holography. Hence, it is crucial to employ more advanced reconstruction algorithms for the single-shot reconstruction, in order to maintain the competitiveness of inline digital holography despite its simpler setup.

In this study, we formulate the reconstruction problem as an inverse problem, where the object field is sought inversely to explain the observed hologram. If we consider an object

with transmittance $\mathbf{x}$ placed in front of the imaging plane at a distance z , the scattered object wave is described using a convolution operation as $\mathbf{u}_o = \mathbf{h}_z \otimes \mathbf{x}$. Here, $\mathbf{h}_z$ represents the free-space propagation kernel and is defined using the Fresnel approximation [1]. Specifically, $\mathbf{h}_z(a, b) = \frac{1}{j\lambda z} e^{j\frac{2\pi}{\lambda} z} e^{j\frac{\pi}{\lambda z}(a^2+b^2)}$ at position (a, b) in the lateral dimensions, with λ denoting the incident wavelength. For the weakly scattering objects, by setting $\mathbf{u}_r = \mathbf{1}$ without loss of generality [14], the forward imaging model in Eq. (1) can be reformulated as

$$\mathbf{y} = 2\text{Re}\{\mathbf{h}_z \otimes \mathbf{x}\} + \mathbf{n} \stackrel{\text{def}}{=} \mathcal{T}(\mathbf{x}) + \mathbf{n}. \quad (2)$$

Here, to simplify the notation, $\mathcal{T}(\cdot)$ denotes the transformation that maps the object field to the hologram intensity. A Gaussian noise term $\mathbf{n} \sim \mathcal{N}(0, \gamma^2 \mathbf{I})$ with variance γ^2 is introduced to account for uncertainties in the physical model including the zero-order frequency terms, sensor thermal noise, optical imperfections, and system misalignment.

For single-shot hologram reconstruction, supervised learning techniques have demonstrated competitive performance by directly mapping recorded holograms to object information [15–21]. However, the effectiveness of these techniques in holographic imaging, particularly in the investigation of biological samples, is hindered by the requirement of a large paired dataset during the training phase. In contrast, unsupervised methods have been investigated, which employ an iterative procedure between the measured and estimated planes while incorporating *a priori* constraints [22–26].

For example, Zhang *et al.* [22] incorporate an explicit prior, i.e., the total variation (TV) regularization during the reconstruction, to enforce the sparsity. However, such an approach requires handcrafted image priors and careful parameter tuning to achieve the desired reconstruction performance. On the other hand, untrained neural network priors (UNNPs) [27] offer distinct advantages by learning implicit prior information directly from the data. This allows for a more adaptable and flexible approach, as UNNPs can effectively combine this learned knowledge with the inherent physics information. Methods like PhysenNet [24] and DeepDIH [25] combine UNNPs with the physics of holography to minimize the reliance on extensive labeled data, enabling the single-shot reconstruction. Galande *et al.* [26] adopt UNNPs empowered with regularization by denoising to mitigate the issue of overfitting the interference-related noise in single-shot measurements. While effective, these methods use randomly initialized convolutional neural networks (CNNs) weights to parameterize the restored image, resulting in a large number of optimization steps due to the complex network architecture and numerous parameters.

Diffusion models are powerful emerging generative models known for capturing complex data distributions and generating high-fidelity samples from noise vectors [28–34]. The rich image priors encompassed in pre-trained diffusion models have shown impressive performance in various vision task [33,34]. However, the potential of diffusion models in digital holography reconstruction has yet to be explored.

In this study, we present the physics-aware diffusion model (PadDH), as a novel approach for unsupervised holographic reconstruction from a single-shot measurement. It leverages the interplay between the physical model and the generative process of the well-trained diffusion model. By incorporating gradient correction and specified initialization, our approach does not require a holographic training dataset, yet it robustly suppresses twin-image noise and accurately reconstructs object information from a single-shot hologram. To the best of our knowledge, this work is the first to explore physics-aware diffusion models in the context of holographic imaging. We compare PadDH with state-of-the-art unsupervised methods that utilize handcrafted priors (i.e., CS [22]) and implicit priors (i.e., DeepDIH [25]) on both synthetic and experimental holographic samples. The results demonstrate the superiority of PadDH in holographic reconstruction, showcasing higher quality results and faster convergence due to the involvement of fewer parameters.

2. Related work on diffusion models

Denoising diffusion models apply diffusion steps to add random noise to data and then learn the reverse process to remove the noise and generate the data samples iteratively. Specifically, a Markov chain is firstly designed to transform any given data distribution $q(\mathbf{x}_0)$ into a simpler prior distribution, typically a standard Gaussian, during the diffusion process. This is represented as $\mathbf{x}_0 \mapsto \mathbf{x}_1 \mapsto \dots \mapsto \mathbf{x}_{T-1} \mapsto \mathbf{x}_T$, where small amounts of Gaussian noise is gradually introduced at each step. By the chain rule and the Markov property, a fixed and factorized posterior distribution can be obtained:

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) = \prod_{t=1}^T q^{(t)}(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad (3)$$

$$q^{(t)}(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}), \quad (4)$$

where $\{\beta_t \in (0, 1)\}_{t=1}^T$ is the variance schedule chosen ahead of model training. To generate new data samples, diffusion models follow a process where they initially sample $\mathbf{x}_T$ from a prior distribution $p(\mathbf{x}_T)$. Subsequently, they iteratively sample $\mathbf{x}_{t-1}$ from the learned Markov chain in the reverse direction. This generative process has a joint distribution defined as

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta^{(t)}(\mathbf{x}_{t-1} | \mathbf{x}_t), \quad (5)$$

where $p_\theta^{(t)}(\mathbf{x}_{t-1} | \mathbf{x}_t)$ is a learned transition kernel with parameters θ . The training objective is to optimize the log-likelihood $\log p_\theta(\mathbf{x}_{0:T})$. This is achieved by minimizing the negative log-likelihood through the utilization of the variational lower bound, which is further simplified into the following loss function

$$L(\theta) = \mathbb{E}_{\mathbf{x}_0, t, \epsilon_t, \sigma_t} [\|\epsilon_\theta(\mathbf{x}_t, t) - \epsilon_t\|_2^2 + \|\sigma_\theta(\mathbf{x}_t, t) - \sigma_t\|_2^2], \quad (6)$$

where $\epsilon_\theta(\mathbf{x}_t, t)$ and $\sigma_\theta(\mathbf{x}_t, t)$ are the output from the trained diffusion network with the noisy $\mathbf{x}_t$ and step t as input. Here, the ϵ_t and σ_t stands for the parameterization of the added noise term at step t .

After training, the diffusion model recovers $\mathbf{x}_0$ by first sampling $\mathbf{x}_T$ from the prior distribution $p_\theta(\mathbf{x}_T)$, and then iteratively restoring $\mathbf{x}_{t-1}$ from the learned transition distribution $p_\theta^{(t)}(\mathbf{x}_{t-1} | \mathbf{x}_t)$ until reaching $\mathbf{x}_0$ via

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{\mathbf{x}}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_\theta(\mathbf{x}_t, t) + \sigma_\theta^2(\mathbf{x}_t, t) \delta, \quad (7)$$

where $\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard Gaussian noise. Here, $\hat{\mathbf{x}}_0$ is the posterior mean of $\mathbf{x}_0$ which is approximated as

$$\hat{\mathbf{x}}_0 \simeq \frac{1}{\sqrt{\bar{\alpha}(t)}} (\mathbf{x}_t + \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{x}_t, t)). \quad (8)$$

The detailed derivations are provided in Appendix A. In light of this, diffusion models excel at capturing intricate data distributions. With a pre-trained diffusion model and employing equations Eq. (7) and Eq. (8), it becomes possible to generate high-fidelity samples from an unstructured noise vector, showcasing impressive unconditional generative modeling performance on images [30–34].

3. Physics-aware diffusion model

In this study, we use the rich neural image prior embedded in an ImageNet-pretrained denoising diffusion probabilistic model (DDPM) [29], considering its diverse and versatile characteristics in the context of different image generation tasks [32–34], and extend its application to the field of holographic imaging. However, this is not straightforward for two main reasons: a) digital inline holographic reconstruction inherently involves a nonlinear inverse problem, and b) the unconditional stochastic sampling step does not provide a means to match the actual measurement. To overcome these challenges, we introduce modifications to the conventional sampling process of DDPM [29] by incorporating gradient correction and specified initialization, which are summarized in Algorithm 1 and visually depicted in Fig. 1. Given that the diffusion priors are pre-trained on a diverse range of natural images, such as ImageNet [35], they inherently impose a real-value constraint. As a result, we constrain the transmittance of our samples to real values within the physical model, with a primary focus on recovering the intensity response of the object. However, we recognize the significance of the phase information as it carries crucial details about the object. In light of this, we explore the potential adaptation of our proposed framework for phase map reconstruction of phase-only objects, which is further discussed in Section 5.

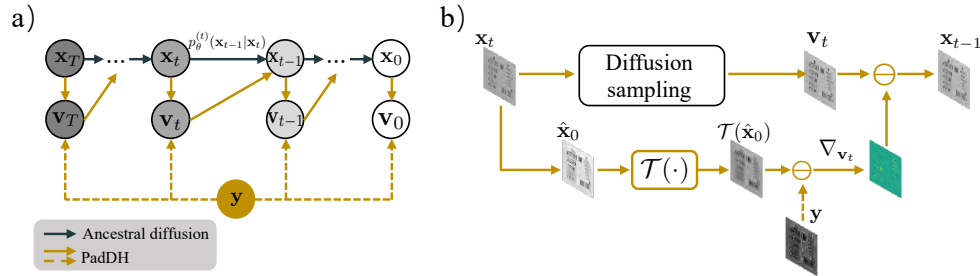


Fig. 1. (a) A comparison between the ancestral diffusion step (depicted in black) and our proposed physics-aware diffusion model (PadDH, represented in orange). It involves an intermediate transformation from x_t to v_t and further transformation into x_{t-1} . (b) A detailed visual depiction of the correction step, illustrating the sequence of transformations from x_t to v_t and then to x_{t-1} .

Algorithm 1. Physics-aware diffusion model for digital holographic reconstruction

Require: DH measurement operator $\mathcal{T}(\cdot)$, hologram y , stepsize ρ , diffusion solver ϵ_θ and σ_θ , number of steps T , variance schedule $\{\beta_t \in (0, 1)\}_{t=1}^T$, initial stage x_T

- 1: **Initialization:** $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$
- 2: **for** $t = T$ to 1 **do**
- 3: $\hat{x}_0 \leftarrow \frac{x_t - \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(x_t, t)}{\sqrt{\bar{\alpha}_t}}$
- 4: $\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$ if $t > 1$, else $\delta = \mathbf{0}$
- 5: $v_t \leftarrow \sqrt{\bar{\alpha}_{t-1}} \hat{x}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_\theta(x_t, t) + \sigma_\theta^2(x_t, t) \delta$ ▷ Diffusion sampling
- 6: $x_{t-1} \leftarrow v_t - \rho \nabla_{v_t} \|y - \mathcal{T}(\hat{x}_0)\|_2^2$ ▷ Gradient correction
- 7: **end for**
- 8: **Return:** x_0

3.1. Gradient correction

To use the physics of digital holographic imaging system as a supervisory signal along the diffusion sampling process, we consider posterior sampling to generate samples that are consistent with the recorded amplitude measurements. Specifically, we maximize the likelihood $p(\mathbf{y} | \mathbf{x}_t)$ of the recorded hologram $\mathbf{y}$ by finding the gradient of it with respect to $\mathbf{x}_t$. However, it is intractable in general. To find a tractable approximation for $p(\mathbf{y} | \mathbf{x}_t)$, we firstly factorize it using the Bayes' theorem:

$$p(\mathbf{y} | \mathbf{x}_t) = \int p(\mathbf{y} | \mathbf{x}_t, \mathbf{x}_0) p(\mathbf{x}_0 | \mathbf{x}_t) d\mathbf{x}_0 = \int p(\mathbf{y} | \mathbf{x}_0) p(\mathbf{x}_0 | \mathbf{x}_t) d\mathbf{x}_0, \quad (9)$$

since that $\mathbf{y}$ and $\mathbf{x}_t$ are conditionally independent on $\mathbf{x}_0$. According to the forward modeling in Eq. (2), $p(\mathbf{y} | \mathbf{x}_0)$ can be expressed as a function of $\mathbf{x}_0$ as

$$g(\mathbf{x}_0) \stackrel{\text{def}}{=} p(\mathbf{y} | \mathbf{x}_0) = \frac{1}{\sqrt{(2\pi)^n \gamma^{2n}}} \exp \left[-\frac{\|\mathbf{y} - \mathcal{T}(\mathbf{x}_0)\|_2^2}{2\gamma^2} \right], \quad (10)$$

where n represents the dimension of the measurement space. It gives

$$p(\mathbf{y} | \mathbf{x}_t) = \int g(\mathbf{x}_0) p(\mathbf{x}_0 | \mathbf{x}_t) d\mathbf{x}_0 \stackrel{\text{def}}{=} \mathbb{E}_{p(\mathbf{x}_0 | \mathbf{x}_t)} g(\mathbf{x}_0). \quad (11)$$

Definition 1 (Jensen's inequality [36]) Let g be a convex or nonconvex function, and let $\mathbf{x}$ be a random vector with distribution $p(\mathbf{x})$. Then $\mathbb{E}_{p(\mathbf{x})} [g(\mathbf{x})] \geq g(\mathbb{E}_{p(\mathbf{x})} [\mathbf{x}])$, where the gap is defined as

$$\mathcal{J} = \mathbb{E}_{p(\mathbf{x})} [g(\mathbf{x})] - g(\mathbb{E}_{p(\mathbf{x})} [\mathbf{x}]). \quad (12)$$

Therefore, based on the Definition 1, the lower bound of $p(\mathbf{y} | \mathbf{x}_t)$ can be derived as

$$p(\mathbf{y} | \mathbf{x}_t) = \mathbb{E}_{p(\mathbf{x}_0 | \mathbf{x}_t)} [g(\mathbf{x}_0)] \geq g(\mathbb{E}_{p(\mathbf{x}_0 | \mathbf{x}_t)} [\mathbf{x}_0]) = p(\mathbf{y} | \hat{\mathbf{x}}_0), \quad (13)$$

where $\hat{\mathbf{x}}_0 \stackrel{\text{def}}{=} \mathbb{E}_{p(\mathbf{x}_0 | \mathbf{x}_t)} [\mathbf{x}_0]$ is the posterior mean of $\mathbf{x}_0$. It implies that an approximation of $p(\mathbf{y} | \mathbf{x}_t)$ can be yield by $p(\mathbf{y} | \hat{\mathbf{x}}_0)$ with the approximation error that is quantified using the Jensen gap, which has a upper bounded in most of the inverse problems [37]. This brings the following approximation

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \mathbf{x}_t) \simeq \nabla_{\mathbf{x}_t} \log p(\mathbf{y} | \hat{\mathbf{x}}_0) = -\frac{1}{\gamma^2} \nabla_{\mathbf{x}_t} \|\mathbf{y} - \mathcal{T}(\hat{\mathbf{x}}_0)\|_2^2, \quad (14)$$

where $\hat{\mathbf{x}}_0$ is a function of $\mathbf{x}_t$ according to Eq. (8). It suggests that taking the gradients $\nabla_{\mathbf{x}_t}$ can be efficiently performed through the differentiable programming [38,39]. Leveraging this capability, an extra correction step based on this gradient term is incorporated into the common ancestral sampling step of the diffusion model with stepsize $\rho = 1/\gamma^2$. To represent this intermediate stage, we introduce a distinct notation of $\mathbf{v}_t$ as a differentiation. Specifically, as shown in the algorithm flow diagram in Fig. 1(a), the ancestral diffusion model goes directly from $\mathbf{x}_t$ to $\mathbf{x}_{t-1}$, while our method introduces the additional gradient correction step. This involves transforming $\mathbf{x}_t$ into $\mathbf{v}_t$ at the intermediate stage before further transforming it into $\mathbf{x}_{t-1}$.

3.2. Specified initialization

Deviating from the conventional diffusion models that start from an unstructured noise vectors, we set the initial point of the generative process as the back-propagated hologram on the object plane, which is

$$\mathbf{x}_T = \mathbf{h}_{-z} \otimes \mathbf{y}, \quad (15)$$

where $\mathbf{h}_{-z}$ is the free-space propagation kernel at the opposite distance $-z$. We reckon that this strategy allows us to leverage the full knowledge of the hologram and initiates the generative

process from a more reasonable and informed starting point. To support this assertion, we conduct comparison experiments with various initialization methods, as detailed in Section 5.

By incorporating the gradient correction step and employing specific initialization, we have effectively merged the physics of the holographic imaging system with the ancestral diffusion model. This integration enables an unsupervised process for a single-shot holographic reconstruction.

4. Experiments

4.1. Optical system

We validate our proposed model using both synthetic and experimental holograms, following the inline holographic imaging system illustrated in Fig. 2. The system begins with a coherent collimated light source module that emits a collimated beam of monochromatic light with a wavelength of λ . The object of interest is positioned in front of the sensor plane at a distance of z . When illuminated, the object scatters the incident light, resulting in an interference pattern when combined with the unscattered reference beam. These holograms are captured by the sensor, which converts the light intensity into electrical signals. The resolution of the sensor, determined by its pixel size, defines the quality of the recorded holograms. Prior to further reconstruction, the holograms undergo pre-processing operations such as cropping and scaling, which resize them to 256×256 pixels and normalize their values to a range of 0 to 1. The size of the input image is governed by the fixed-size constraint inherent in the pre-trained diffusion model utilized within PadDH. This constraint allows us to leverage the power of pre-existing models without the necessity of re-training or fine-tuning. However, we acknowledge that it is indeed possible to accommodate larger image sizes by utilizing alternative pre-existing diffusion models in PadDH that are trained on such dimensions. The system configurations for both simulation and experiment are listed in Table 1. The reconstruction performance is compared against other state-of-the-art unsupervised methods that utilize handcrafted priors (i.e., CS [22]) and implicit priors (i.e., DeepDIH [25]).

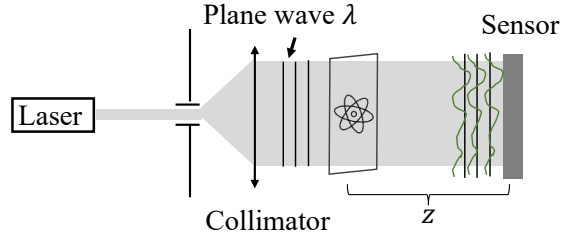


Fig. 2. Schematic setup for digital inline holography used in simulations and experiments.

Table 1. System configurations for simulation and experimental setup.
The rescaled pixel size is derived from the original hologram, which undergoes cropping and scaling operations to match the network input size of 256×256 .

System configurations	Simulation		Experiment	
	USAF	Cell	USAF	Convallaria
Wavelength λ (nm)	532.0	532.0	632.8	532.0
Distance z (cm)	0.1	0.1	2.28	3.33
Pixel size (μm)	1.12	1.12	4.80	3.45
Rescaled pixel size (μm)	1.12	1.12	13.5	13.5

4.2. Evaluation with simulated holograms

To validate our algorithms, we initially conduct simulations. To explore the generalization ability of our developed framework across different sample spaces, we conduct tests on a set of synthetic holograms generated from a standard resolution USAF target and open-source biological cell images [40]. The samples are employed as amplitude objects within an inline digital holography system, as described in Section 4.1. Both types of samples adhere to the same configuration, which is summarized in Table 1.

Figure 3 illustrates the reconstructed results of the standard USAF target, alongside the ground truth. Notably, the BP result exhibits significant twin-image contamination. In contrast, both the CS [22] and DeepDIH [25] algorithms, as well as our proposed method (PadDH), effectively mitigate this disturbance. To facilitate a comprehensive visual comparison, we employ a Canny edge detector to extract the edge matrix from each reconstructed output. To ensure fairness, all edge matrix images are calculated using the same threshold by the operator. Two regions of interest (ROI) with different resolution scales are then enlarged. Additionally, cross-section plots along the selected line are generated for the chosen area. Evidently, PadDH demonstrates superior alignment between the extracted result and the ground truth. The edges appear sharper, while background noise is significantly suppressed. These observations strongly indicate that the PadDH algorithm outperforms the BP, CS [22], and DeepDIH [25] methods in effectively suppressing twin-image noise.

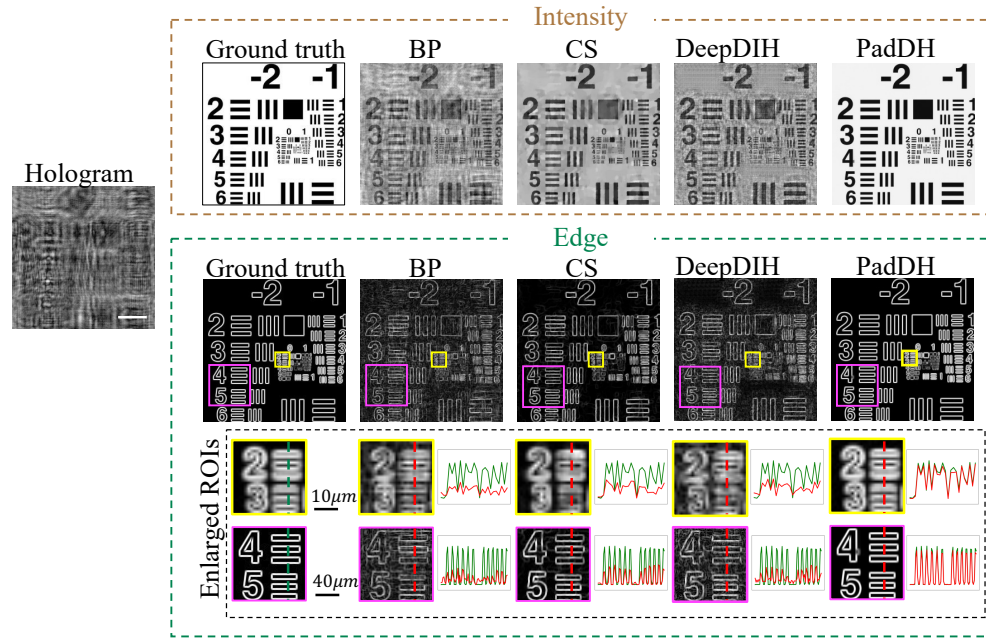


Fig. 3. Reconstruction results of USAF target with comparison against BP, CS [22] and DeepDIH [25]. The reconstructed intensity images are processed by Canny edge detector to generate edge matrix for a better visual comparison. Enlarged regions of interest (ROIs) corresponding to different resolution scales are shown below each result. Scale bars for the entire field of view (FOV) and the magnified ROIs: 60 μ , 40 μ m and 10 μ m, respectively.

The generalization ability of our proposed framework across different sample domains is verified using the biological cell samples [40], which exhibit distinct semantic content compared to natural images. To thoroughly evaluate our framework, we select a diverse set of 500 cell samples, each containing various subcellular structures. The reconstruction results, depicted in

Fig. 4 with a cool-warm colormap, effectively capture the internal structure of the cells, including the nucleus, apparatus, and fibers. The enlarged area demonstrates that our method successfully retrieves the in-focus object, preserving fine features of the internal structure of the cells, and effectively eliminates the unwanted twin-image effects caused by the out-of-focus regions.

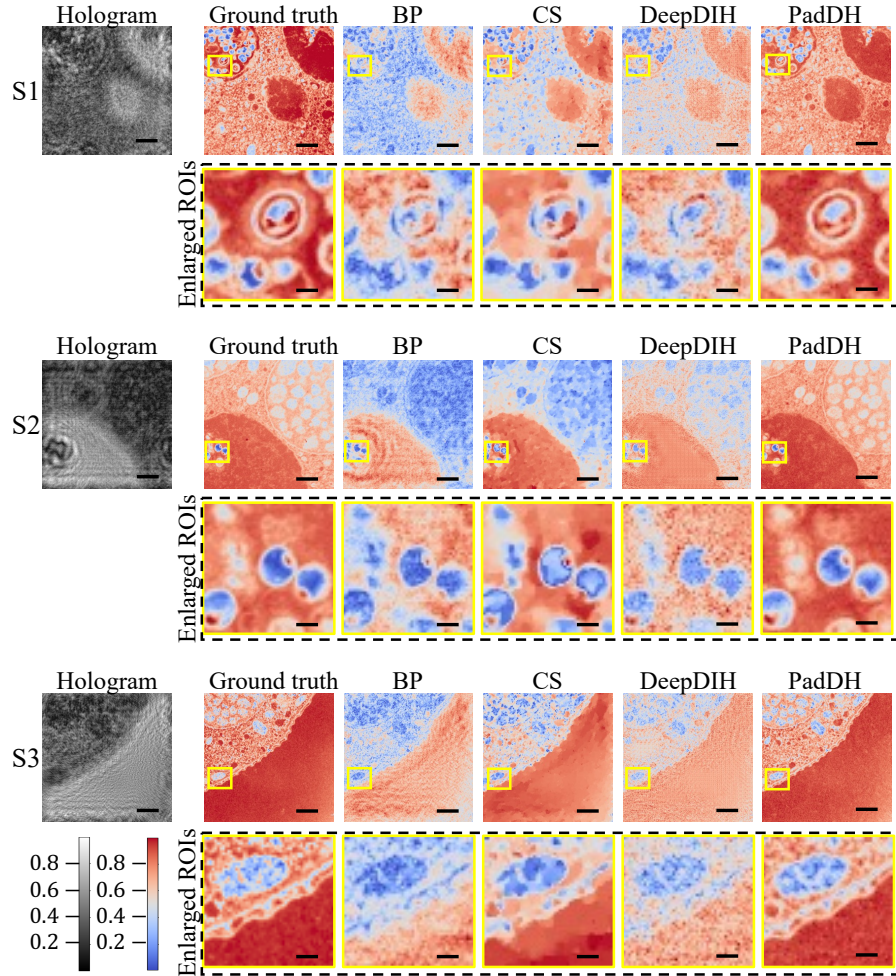


Fig. 4. Reconstruction results of cell samples (S1, S2, and S3) generated from open-source biological cell images [40]. For each case, the first row displays the synthetic hologram, the ground truth and the reconstruction results obtained using the conventional BP, CS [22], DeepDIH [25] and our proposed PadDH subsequently in the entire field of view (FOV). The second row shows enlarged regions of interest (ROIs), highlighting the consistent ability of our proposed method to resolve fine details of the cell structure. Scale bars for the entire field of view (FOV) and the magnified ROIs: 60 μm and 10 μm , respectively.

To quantitatively evaluate the reconstruction results, we employ widely recognized metrics i.e., the peak signal-to-noise ratio (PSNR), the structural similarity index measure (SSIM), and the root mean square error (RMSE) to assess the similarity between the reconstructed images and their corresponding ground truth counterparts. While these metrics have long been utilized for evaluating image quality, it is important to note that certain types of degradation may not be accurately represented in terms of perceived image quality. To provide a more comprehensive assessment, we incorporate three additional mathematically defined image quality

assessment (IQA) models, i.e., the feature-based similarity index (FSIM) [41], the universal image quality index (UIQ) [42], and the normalized mutual information (NMI) [43]. By including these additional metrics, we aim to capture a broader perspective on the performance of the reconstruction methods. The averaged evaluation metrics are summarized in Table 2. It demonstrates significant performance improvements achieved by our method across multiple evaluation metrics, highlighting its superior efficacy in digital holography reconstruction. These results also suggest the generalization capabilities of our method across different problem domains, as evidenced by its consistent performance on both the USAF target and biological cell samples.

Table 2. Quantitative evaluation of different reconstruction methods. The arrow direction beside each metric indicates better reconstruction quality.

Samples	Methods	Metrics					
		PSNR(dB) ↑	SSIM ↑	RMSE ↓	FSIM ↑	UIQ ↑	NMI ↑
USAF	BP	8.92	0.34	0.36	0.38	0.49	1.09
	CS [22]	10.89	0.52	0.46	0.61	0.44	1.13
	DeepDIH [25]	9.14	0.43	0.35	0.42	0.57	1.10
	PadDH	25.47	0.96	0.05	0.75	0.99	1.35
cells	BP	10.08	0.40	0.31	0.50	0.44	1.08
	CS [22]	9.96	0.53	0.32	0.59	0.56	1.11
	DeepDIH [25]	15.48	0.50	0.17	0.58	0.67	1.09
	PadDH	23.19	0.67	0.07	0.67	0.81	1.14

4.3. Evaluation with real holograms

For our experimental validation, a resolution target is firstly employed as a transmissive sample. As depicted in Fig. 5(a), our experimental setup involves a laser module (HNL100L, Thorlabs) that generates a 10 mW single wavelength (632.8 nm) light source. A digital camera, specifically the NV-GE134GM-T model from MindVision, equipped with a CMOS sensor featuring a pixel size of 4.8 μm , is employed to capture the holograms. To ensure accurate hologram recording, the collimated and expanded light passes through a beam expander (BE10M-A, Thorlabs) before interacting with the sample. A neutral density filter (ND) controls the light intensity to prevent sensor overexposure. The sample, positioned in the light path, propagates the hologram onto the sensor plane. The results are demonstrated in Fig. 5(b), where it becomes evident that PadDH outperforms the others in terms of noise suppression. Notably, the intensity profiles along the selected line are plotted in Fig. 5(c) with different colors to represent reconstruction results from BP, CS [22], DeepDIH [25] and our method. Upon closer inspection of the zoomed-in plots, PadDH stands out by effectively mitigating disturbances caused by twin-image noise and outperforming others through sharper edges and a more consistent background. It highlights the superior performance of PadDH, demonstrating its distinctiveness and effectiveness compared to alternative methods.

Moreover, to assess the potential of our method in biomedical applications, we conduct experiments using experimental data from a convallaria sample collected under the same optical system setup [39]. This experiment also aims to demonstrate the generalizability of our approach across different system configurations, which are specified in Table 1. By observing the enlarged area ROI 3, our approach keeps its superior performance by preserving high-quality details to a significant extent. Since no ground-truth data is available for comparison, we quantify the effectiveness of our method's twin-image removal capability using no-reference image quality assessment algorithms. Specifically, we calculate the total variation (TV), where smaller values indicate more effective suppression of twin-image artifacts and a smoother background. The TV

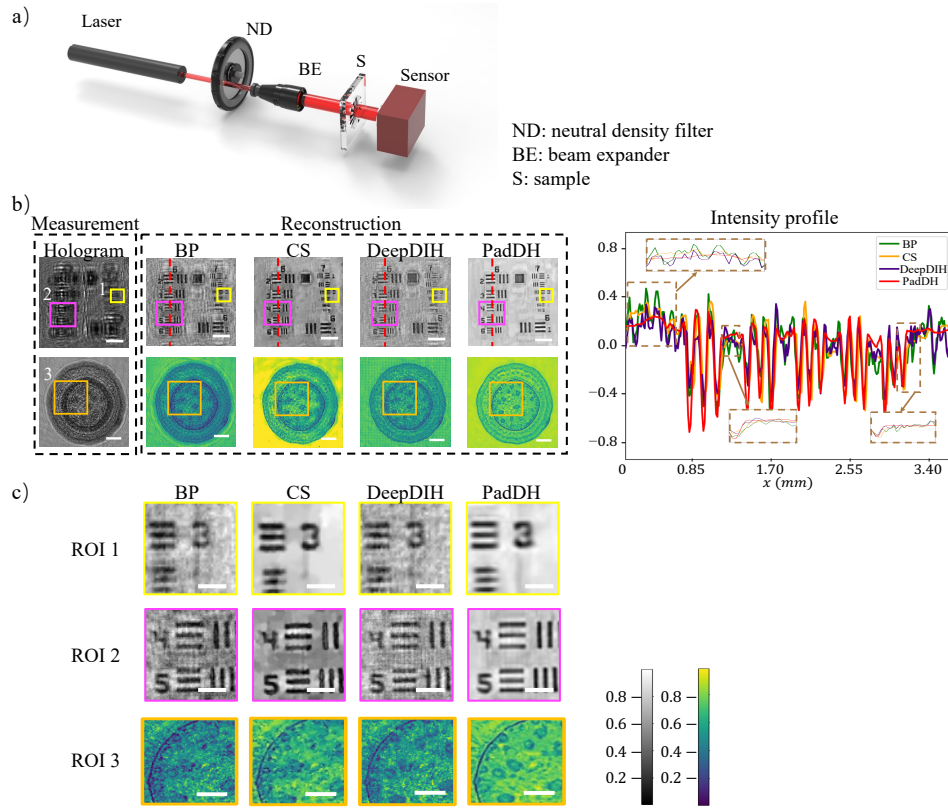


Fig. 5. Experimental verification. (a) Experimental setup of the optical system. (b) Reconstruction results of our proposed method on two samples, i.e. a resolution target and a convallaria sample. Intensity profiles along the dashed red line of interest for each method are plotted. Selected ranges are zoomed in dashed rectangular boxes. (c) Three ROIs in (b) are selected and enlarged to reflect the reconstruction details. Scale bars for the entire FOV and three enlarged ROIs: 0.7 mm, 0.2 mm, 0.3 mm and 0.4 mm, respectively.

values calculated for the BP, CS [22], DeepDIH [25], and PadDH methods are 0.157, 0.082, 0.133 and 0.058, respectively. Additionally, the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [44] is used, which employs natural scene statistics to assess the quality of an image. It relies on spatial natural scene statistics models of locally normalized luminance coefficients and pairwise products of these coefficients in the spatial domain. The BRISQUE [44] scale, which ranges from 0 to 100 (with lower values indicating better subjective quality), yields scores of 38.5, 17.8, 29.4 and 4.79 for the BP, CS [22], DeepDIH [25], and PadDH, respectively. These results further reinforce the superiority of our reconstructed images in terms of human judgments of image quality. These comparative results highlight the exceptional performance of PadDH, demonstrating its distinctiveness and effectiveness compared to alternative methods.

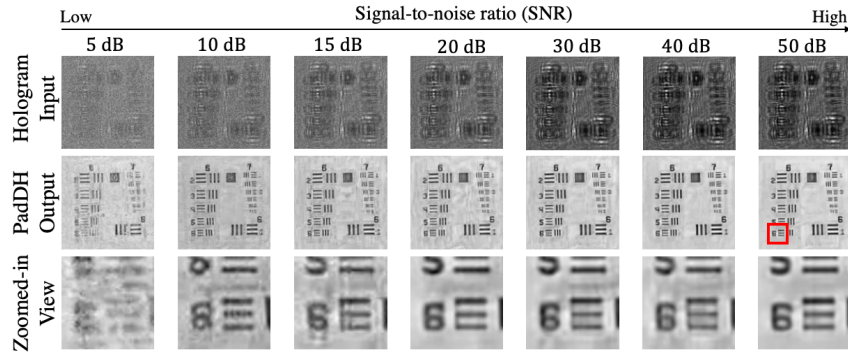
4.4. Computational cost and robustness investigation

In terms of computational expense, CS [22] converges faster since it does not involve any trainable parameters. However, the reconstruction quality of this method is subject to the influence of domain-specific priors, leading to performance variations across different problem domains. In contrast, both DeepDIH [25] and our method (PadDH) incorporate implicit priors, but PadDH significantly reduces the number of trainable parameters, from 21 million to 0.06 million. This

reduction in trainable parameters leads to a substantial decrease in computational expense while maintaining comparable reconstruction quality.

For real-world imaging scenarios with low signal-to-noise ratio (SNR), it is crucial to assess the robustness of the computational reconstruction algorithm under different noise levels. We introduce Gaussian noise with varying degrees of standard deviation to the experimental USAF target, resulting in a reduction of SNR from 50 dB to 5 dB. As depicted in Fig. 6(a), the interference patterns gradually deteriorate and become contaminated as the SNR decreases. However, PadDH consistently demonstrates its effectiveness across different noise levels. This robustness can be attributed, in part, to the inherent noise removal capabilities of diffusion models. Nevertheless, when the SNR of the holograms decreases to 5 dB, the holograms become excessively noisy, rendering them unsuitable for PadDH to reconstruct detailed features. This limitation arises due to the high level of corruption in the interference patterns under such noisy conditions, which hampers the provision of sufficient information for an accurate reconstruction.

(a)



(b)

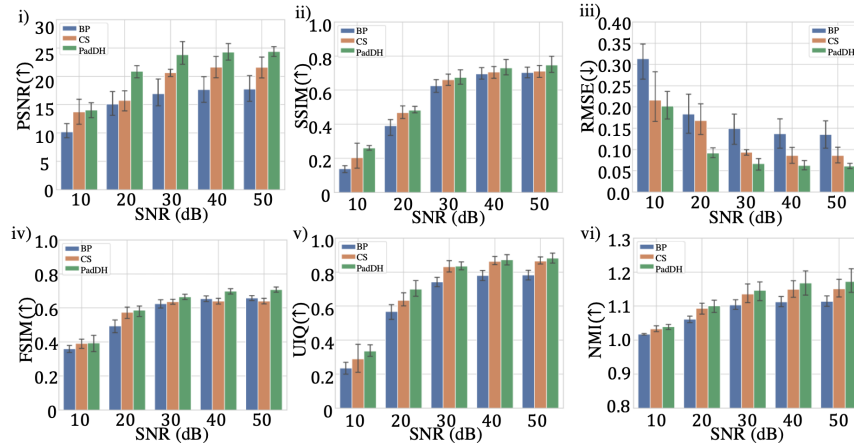


Fig. 6. Robustness verification across different signal-to-noise ratios (SNR). (a) *First row:* Input holograms of USAF target arranged in ascending order of SNR, acquired with an inline holographic system. *Second row:* USAF target reconstructed using PadDH. *Third row:* Zoomed-in views of the reconstructed region (highlighted by the red box). (b) Quantitative comparison on noisy cell holograms with varying noise levels. Six different metrics are utilized and presented as bar plots, indicating the means and 95% confidence intervals.

To provide a quantitative comparison between PadDH and other methods, we conduct tests on corrupted synthetic cell holograms where the ground truths are available. We generate six groups of noisy cell holograms, each containing 100 data samples, with varying SNRs. The reconstruction results of BP, CS [22], and PadDH are compared against the ground truth, and evaluation metrics are calculated and visualized using barplots in Fig. 6(b). The results clearly indicate that PadDH exhibits greater robustness to noisy data compared to both BP and CS [22]. This is evidenced by its consistent performance across multiple evaluation scenarios, as demonstrated by the barplots. PadDH outperforms the other methods, providing more accurate reconstructions even in the presence of significant noise.

5. Discussion and limitations

5.1. Investigation on different initializations

The influence of the specified initialization in PadDH is investigated by comparing the results obtained from two different initializations: randomized initialization commonly used in conventional diffusion models, and the specified initial point used in PadDH. We conduct the comparison using the experimental hologram from the convallaria sample, and the results are shown in Fig. 7(a). As depicted in Fig. 7(a), with randomized initialization, the reconstruction output is unstable and exhibits undesired artifacts. In contrast, our model, which utilizes the specified initialization, can effectively recover the details of the object with improved accuracy. This comparison provides evidence of the advantages of our specified initialization in merging the imaging physics of the holographic imaging system into ancestral diffusion models, leading to enhanced reconstruction results.

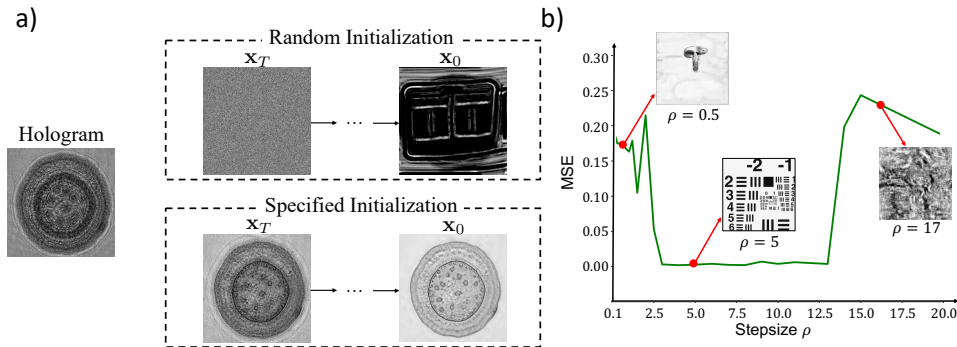


Fig. 7. Impact of initialization and stepsize on reconstruction performance. (a) The comparison of reconstruction results using different initialization x_T . (b) The mean-square-error (MSE) between the ground truth and reconstruction results using different step sizes ranging from 0.1 to 20.

5.2. Investigation on stepsize ρ

The choice of hyperparameter setting for the step size ρ influences the stability of the algorithm. This parameter is closely related to the noise value, indicating the degree of influence enforced by the physics information on the generative process of the diffusion model. To investigate its impact, we conduct an ablation study where we varied the value of the step size and examined its effect on the reconstruction performance. Specifically, we calculate the mean-square-error (MSE) between the ground truth and the reconstruction results obtained using different step sizes, ranging from 0.1 to 20. As illustrated in Fig. 7(b), we observed that if the physics information is under-weighted, the posterior becomes over-biased toward the diffusion

priors, potentially leading to misleading data. Conversely, if the measurements are over-weighted, the samples may collapse onto a subspace that does not align with the diffusion priors. These observations highlight the importance of tuning the step size for stability.

However, within a stable range of step sizes, we find that the reconstruction performance remained consistent, suggesting a relative safety buffer range for tuning. In our simulations, we empirically optimize the step size to ensure sampling stability and reliable reconstructions. We then apply the same optimized value directly to the experimental samples, successfully reconstructing the object information with good stability and performance. It is important to note that the specific value of the step size may vary depending on the experimental setup and dataset used. Further optimization of the step size holds the potential for enhancing reconstruction results, which is beyond the scope of this study.

5.3. Investigation on phase recovery

In this study, we investigate the application of the diffusion model as an image prior for unsupervised single-shot holographic imaging. Our method integrates the image formation physics into the pre-existing diffusion model, such as DDPM [29], without the need for re-training or fine-tuning that requires adequate experimental holograms. Given the inherent real-value constraint present in these diffusion priors, our primary emphasis is on recovering the intensity response of the object by assuming the real-valued transmittance of our samples within the physical model. However, the phase information recovery of the object is another important application in holographic imaging. With this in mind, we investigate the adaptiveness of PadDH in phase recovery. In Appendix B, we provide a detailed account of how PadDH can be adapted for phase-only object reconstruction. Specifically, we employ the same PadDH architecture while modifying the object model with its real-valued phase map. We refer to this variation as PadPhase for phase-only object reconstruction. Through our experiments, we observe that PadPhase has desirable reconstruction for the given phase recovery problem, capturing most of the high-frequency details. This suggests the generalization ability of our proposed framework in solving reconstruction problems with different imaging modalities.

However, it is important to consider that in practical applications, the presence of phase-only substances is limited, as most tissues and living cells exhibit absorption in the visible spectrum and additional optical effects such as refraction. Consequently, to fully unlock the potential of our method in holographic imaging, the recovery of the complex transmittance, encompassing both amplitude and phase simultaneously, becomes crucial. Addressing this challenge will be a significant focus of our future research endeavors.

6. Conclusion

In this paper, we present a novel physics-aware diffusion model (PadDH), designed specifically for unsupervised single-shot holographic imaging. This work introduces the use of diffusion models in holographic reconstruction by combining the physical information and the effective image prior embedded in a pre-trained diffusion model. Comparative evaluations with state-of-the-art unsupervised methods through comprehensive experiments highlight the advantages of PadDH. The results demonstrate its superiority in holographic reconstruction, showcasing improved quality results and faster convergence. Additionally, our investigation highlights the robustness of PadDH across various noise levels, making it highly suitable for real-world applications characterized by low signal-to-noise ratios. This advancement holds crucial applications for compact and portable inline holographic imaging systems, which has huge impact for on-site or field investigations. Besides, this research opens up new possibilities for enhancing scientific imaging by leveraging the powerful prior knowledge encoded in pre-trained diffusion models. Future work can further explore and optimize PadDH to unlock its full potential for a wide range of scientific imaging applications.

Appendix A: Derivation of loss function (Eq. (5)) and update process (Eq. (6)) in ancestral diffusion models

In this study, we use the ImageNet-pretrained DDPM [29] as the effective diffusion prior considering its diverse and versatile characteristics in the context of different image generation tasks. The optimization of DDPM is realized by minimising the variational bound on the negative log-likelihood through

$$\mathbb{E}_{q(\mathbf{x}_{0:T})} [-\log p_{\theta}(\mathbf{x}_{0:T})] \leq \mathbb{E}_{q(\mathbf{x}_{0:T})} \left[-\log \frac{p_{\theta}(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)} \right], \quad (16)$$

$$= \mathbb{E}_{q(\mathbf{x}_{0:T})} \left[-\log p(\mathbf{x}_T) - \sum_{t=1}^T \log \frac{p_{\theta}^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t)}{q^{(t)}(\mathbf{x}_t|\mathbf{x}_{t-1})} \right]. \quad (17)$$

The learning objective can be further formulated to minimize the Kullback–Leibler (KL) divergence between the trackable conditioned posteriors $q^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ and $p_{\theta}^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t)$ (please refer to [29] for detailed derivations).

Based on the Gaussian transitions defined in Eq. (3) and Bayes theorem, $q^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ is a Gaussian distribution with the mean and variance calculated as

$$\begin{aligned} \boldsymbol{\mu}_t &= \frac{\sqrt{\alpha_t}(1 - \bar{\alpha}_{t-1})\mathbf{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1 - \alpha_t)\mathbf{x}_0}{1 - \bar{\alpha}_t}, \\ \boldsymbol{\Sigma}_t &= \frac{(1 - \alpha_t)(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} \mathbf{I} = \sigma_t^2 \mathbf{I}, \end{aligned} \quad (18)$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ for notation simplicity. There are several options available for parameterizing the learnable transition kernel $p_{\theta}^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t)$. The most obvious option is the Gaussian distribution

$$p_{\theta}^{(t)}(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t), \sigma_{\theta}^2(\mathbf{x}_t, t) \mathbf{I}), \quad (19)$$

where $\boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t)$ and $\sigma_{\theta}(\mathbf{x}_t, t)$ is the output from the learned deep neural network with the noisy $\mathbf{x}_t$ and step t as input. It is worth noting that, instead of training different networks for T steps, the parameters θ can be shared across different steps by the Transformer sinusoidal position embedding [45].

Therefore, the training objective is to predict a pair of $\boldsymbol{\mu}_{\theta}(\mathbf{x}_t, t)$ and $\sigma_{\theta}(\mathbf{x}_t, t)$ given the input $\mathbf{x}_t$ and step t , aiming to closely approximate the values $\boldsymbol{\mu}_t$ and σ_t . To enhance both the quality and efficiency of sampling, Ho et al [29] discovered the advantages of predicting the added noise term at step t instead. This approach is founded on the following linear transformation between the noisy $\mathbf{x}_t$ and clean $\mathbf{x}_0$, taking into account the properties of Gaussian transitions as described in Eq. (3)

$$\mathbf{x}_t = \sqrt{\bar{\alpha}(t)}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}(t)}\boldsymbol{\epsilon}_t, \quad (20)$$

where $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard Gaussian noise. Therefore, the total loss is defined as

$$L(\theta) = \mathbb{E}_{\mathbf{x}_0 \sim q(\mathbf{x}_0), t \sim [0, T], \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} [\|\boldsymbol{\epsilon}_{\theta}(\mathbf{x}_t, t) - \boldsymbol{\epsilon}_t\|_2^2 + \|\sigma_{\theta}(\mathbf{x}_t, t) - \sigma_t\|_2^2], \quad (21)$$

where $\boldsymbol{\epsilon}_{\theta}(\mathbf{x}_t, t)$ and $\sigma_{\theta}(\mathbf{x}_t, t)$ are the output from the learned deep neural network with the noisy $\mathbf{x}_t$ and step t as input.

After training, given the input $\mathbf{x}_t$ at step t , the learned neural network can generate a pair of $\epsilon_\theta(\mathbf{x}_t, t)$ and $\sigma_\theta(\mathbf{x}_t, t)$. This gives the following iterative update in the generative process

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_\theta^2(\mathbf{x}_t, t) \delta, \quad (22)$$

where $\delta \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard Gaussian noise. The linear transformation between the noisy $\mathbf{x}_t$ and clean $\mathbf{x}_0$ in the diffusion process described in Eq. (20) can be re-written by means of

$$\mathbf{x}_0 = \frac{1}{\sqrt{\bar{\alpha}(t)}} \left(\mathbf{x}_t + \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{x}_t, t) \right). \quad (23)$$

Therefore, the posterior mean of $\mathbf{x}_0$ can be approximated as

$$\hat{\mathbf{x}}_0 \simeq \frac{1}{\sqrt{\bar{\alpha}(t)}} \left(\mathbf{x}_t + \sqrt{1 - \bar{\alpha}_t} \epsilon_\theta(\mathbf{x}_t, t) \right), \quad (24)$$

where $\epsilon_\theta(\mathbf{x}_t, t)$ is the output from the learned network. This finally rewrites the generative process in Eq. (22) into

$$\mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}} \hat{\mathbf{x}}_0 + \sqrt{1 - \bar{\alpha}_{t-1}} \epsilon_\theta(\mathbf{x}_t, t) + \sigma_\theta^2(\mathbf{x}_t, t) \delta, \quad (25)$$

which is the Eq. (7) in the main text.

Appendix B: Investigation on phase recovery

Here, we present an investigation into the application of PadDH for phase-only object reconstruction. Considering the phase-only prior of the object, we modify the forward imaging model in Eq. (2) from

$$\mathbf{y} = 2\text{Re}\{\mathbf{h}_z \otimes \mathbf{x}\} + \mathbf{n}, \quad (26)$$

into

$$\mathbf{y} = 2\text{Re}\{\mathbf{h}_z \otimes e^{j\mathbf{x}}\} + \mathbf{n}, \quad (27)$$

where $\mathbf{x}$ stands for the phase map of the object that is real-valued. We termed this variation of our proposed framework as PadPhase, which highlights its use in phase reconstruction. To assess its performance, we conduct experiments using a standard USAF target and cell samples to evaluate its resilience across different sample spaces. As shown in Fig. 8, PadPhase consistently recovers the phase information with remarkable consistency across diverse samples, showcasing the adaptiveness of our proposed framework.

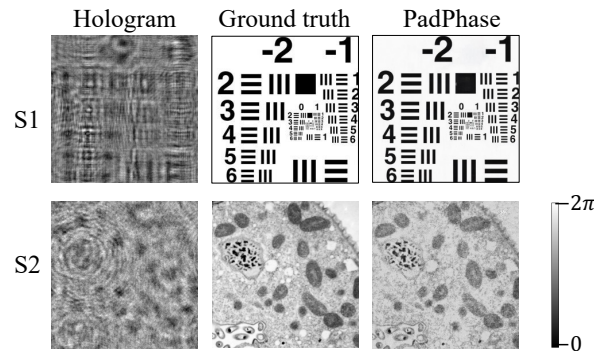


Fig. 8. Phase reconstruction by PadPhase for two synthetic holograms of USAF target (S1) and cell sample (S2). The sequence includes the synthetic holograms, the ground-truth phase maps, and the corresponding reconstruction results.

Funding. Research Grants Council of Hong Kong (GRF 17201620, GRF 17201822, RIF 7003-21).

Acknowledgment. The authors would like to express sincere gratitude to Dr. Ni Chen for providing access to the convallaria data used in this study.

Disclosures. The authors declare no conflicts of interest.

Data availability. All the data and codes are available at [46].

References

1. J. W. Goodman, *Introduction to Fourier Optics* (W.H. Freeman & Company Ltd, 2017), 4th ed.
2. G. Popescu, *Quantitative Phase Imaging of Cells and Tissues* (McGraw-Hill Education, 2011).
3. Y. Zhang, Y. Zhu, and E. Y. Lam, "Holographic 3D particle reconstruction using a one-stage network," *Appl. Opt.* **61**(5), B111–B120 (2022).
4. N. Chen, C. Wang, and W. Heidrich, "Holographic 3D particle imaging with model-based deep network," *IEEE Trans. Comput. Imaging* **7**, 288–296 (2021).
5. O. Mudanyali, D. Tseng, C. Oh, *et al.*, "Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications," *Lab Chip* **10**(11), 1417–1428 (2010).
6. T. Liu, Y. Li, H. C. Koydemir, *et al.*, "Rapid and stain-free quantification of viral plaque via lens-free holography and deep learning," *Nat. Biomed. Eng.* **7**(8), 1040–1052 (2023).
7. Y. Zhu, C. H. Yeung, and E. Y. Lam, "Microplastic pollution monitoring with holographic classification and deep learning," *J. Phys. Photonics* **3**(2), 024013 (2021).
8. Y. Zhu, H. K. A. Lo, C. H. Yeung, *et al.*, "Microplastic pollution assessment with digital holography and zero-shot learning," *APL Photonics* **7**(7), 1 (2022).
9. C. Shen, M. Liang, A. Pan, *et al.*, "Non-iterative complex wave-field reconstruction based on Kramers-Kronig relations," *Photonics Res.* **9**(6), 1003–1012 (2021).
10. A. Greenbaum and A. Ozcan, "Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy," *Opt. Express* **20**(3), 3129–3143 (2012).
11. W. Luo, Y. Zhang, Z. Göröcs, *et al.*, "Propagation phasor approach for holographic image reconstruction," *Sci. Rep.* **6**(1), 22738 (2016).
12. H. Zhang, T. Stangner, K. Wiklund, *et al.*, "Object plane detection and phase retrieval from single-shot holograms using multi-wavelength in-line holography," *Appl. Opt.* **57**(33), 9855–9862 (2018).
13. C. Shen, X. Bao, J. Tan, *et al.*, "Two noise-robust axial scanning multi-image phase retrieval algorithms based on pautra criterion and smoothness constraint," *Opt. Express* **25**(14), 16235–16249 (2017).
14. D. J. Brady, K. Choi, D. L. Marks, *et al.*, "Compressive holography," *Opt. Express* **17**(15), 13040–13049 (2009).
15. H. Luo, J. Xu, L. Zhong, *et al.*, "Diffraction-Net: a robust single-shot holography for multi-distance lensless imaging," *Opt. Express* **30**(23), 41724–41740 (2022).
16. H. Wang, M. Lyu, and G. Situ, "eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction," *Opt. Express* **26**(18), 22603–22614 (2018).
17. K. Wang, L. Song, C. Wang, *et al.*, "On the use of deep learning for phase recovery," *Light: Sci. Appl.* **13**(1), 4 (2024).
18. K. Wang, J. Dou, Q. Kemao, *et al.*, "Y-Net: a one-to-two deep learning framework for digital holographic reconstruction," *Opt. Lett.* **44**(19), 4765–4768 (2019).
19. Y. Rivenson, Y. Zhang, H. Günaydin, *et al.*, "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light: Sci. Appl.* **7**(2), 17141 (2017).
20. M. Rogalski, P. Arcab, L. Stanaszek, *et al.*, "Physics-driven universal twin-image removal network for digital in-line holographic microscopy," *Opt. Express* **32**(1), 742–761 (2024).
21. C. Bai, T. Peng, J. Min, *et al.*, "Dual-wavelength in-line digital holography with untrained deep neural networks," *Photonics Res.* **9**(12), 2501–2510 (2021).
22. W. Zhang, L. Cao, D. J. Brady, *et al.*, "Twin-image-free holography: a compressive sensing approach," *Phys. Rev. Lett.* **121**(9), 093902 (2018).
23. Y. Rivenson, Y. Wu, H. Wang, *et al.*, "Sparsity-based multi-height phase recovery in holographic microscopy," *Sci. Rep.* **6**(1), 37862 (2016).
24. F. Wang, Y. Bian, H. Wang, *et al.*, "Phase imaging with an untrained neural network," *Light: Sci. Appl.* **9**(1), 77 (2020).
25. H. Li, X. Chen, Z. Chi, *et al.*, "Deep DIH: Single-shot digital in-line holography reconstruction by deep learning," *IEEE Access* **8**, 202648–202659 (2020).
26. A. S. Galande, V. Thapa, H. P. R. Gurram, *et al.*, "Untrained deep network powered with explicit denoiser for phase recovery in inline holography," *Appl. Phys. Lett.* **122**(13), 1 (2023).
27. A. Qayyum, I. Iah, F. Shamshad, *et al.*, "Untrained neural network priors for inverse imaging problems: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
28. X. Li, Y. Ren, X. Jin, *et al.*, "Diffusion models for image restoration and enhancement—a comprehensive survey," *arXiv arXiv:2308.09388* (2023).
29. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems* **33**, 6840–6851 (2020).

30. P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in Neural Information Processing Systems* **34**, 8780–8794 (2021).
31. L. Yang, Z. Zhang, Y. Song, *et al.*, "Diffusion models: A comprehensive survey of methods and applications," *arXiv* arXiv:2209.00796 (2022).
32. Y. Song, J. Sohl-Dickstein, D. P. Kingma, *et al.*, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, (2020).
33. J. Wang, Z. Yue, S. Zhou, *et al.*, "Exploiting diffusion prior for real-world image super-resolution," *arXiv* arXiv:2305.07015 (2023).
34. B. Fei, Z. Lyu, L. Pan, *et al.*, "Generative diffusion prior for unified image restoration and enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023), pp. 9935–9946.
35. J. Deng, W. Dong, R. Socher, *et al.*, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, (2009), pp. 248–255.
36. M. DeGroot and M. Schervish, *Probability and Statistics*, Pearson custom library (Pearson Education, 2013).
37. X. Gao, M. Sitharam, and A. E. Roitberg, "Bounds on the Jensen Gap, and implications for mean-concentrated distributions," *The Australian Journal of Mathematical Analysis and Applications* **16**(2), 1–16 (2019).
38. N. Chen, L. Cao, T.-C. Poon, *et al.*, "Differentiable imaging: A new tool for computational optical imaging," *Adv. Phys. Res.* **2**(6), 2200118 (2023).
39. N. Chen, C. Wang, and W. Heidrich, " $\partial\mathbf{H}$: Differentiable holography," *Laser Photonics Rev.* **17**(9), 1 (2023).
40. R. Allen, "Vorticella convallaria, cell by organism, eukaryotic cell, eukaryotic protist, ciliated protist," Dataset CIL:39327, Cell Image Library (2012). <https://doi.org/doi:10.7295/W9CIL39327>.
41. L. Zhang, L. Zhang, X. Mou, *et al.*, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. on Image Process.* **20**(8), 2378–2386 (2011).
42. Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
43. C. Studholme, D. L. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3D medical image alignment," *Pattern Recognit.* **32**(1), 71–86 (1999).
44. A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. on Image Process.* **21**(12), 4695–4708 (2012).
45. A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems* **30**, 1 (2017).
46. Y. Zhang, X. Liu, and E. Y. Lam, "PadDH:Physics-aware diffusion model," Github (2023). <https://github.com/yp000925/PadDH>