Research Article Vol. 33, No. 3 / 10 Feb 2025 / Optics Express 5932
Optics EXPRESS

Robust holographic imaging for real-world applications with joint optimization

YUNPING ZHANG AND EDMUND Y. LAM*

Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China
*elam@eee.hku.hk

Abstract: Digital inline holography offers a compact, lensless imaging solution, but its practical deployment is often hindered by the need for precise system alignment and calibration, particularly regarding propagation distance. This work presents J-Net, a robust, untrained neural network that significantly mitigates these limitations. J-Net eliminates the need for prior knowledge or calibration of the propagation distance by simultaneously reconstructing both the complex-valued object magnitude and the propagation distance from a single hologram. This inherent robustness to distance variations makes J-Net highly practical for real-world applications where precise system control is difficult or impossible. Experimental results demonstrate high-quality amplitude and phase reconstruction even under mismatched distance conditions, showcasing J-Net's potential to enable robust deployment of holographic imaging across diverse fields.

1. Introduction

Holographic imaging serves as a powerful technique enabling the retrieval of complex-valued object wavefields from intensity-only measurements. This is achieved by illuminating the object with coherent light and recording the interference pattern (hologram) formed by the superposition of the object's scattered wave and the unscattered reference wave. Digital reconstruction algorithms are then used to extract the complex-valued object information (i.e. amplitude and phase), which has broad applications across fields like biomedical sciences, physical sciences, and engineering [1–3]. Basing on their involvement with the physical model, existing reconstruction methods can be divided into three categories as summarized in Table 1.

Traditional model-based methods [4–6] offer interpretability and consistency with minimal data but can be slower and require meticulous manual tuning. Deep learning methods [17–20] excel in performance and adaptability but demand substantial data and computational resources. As a middle ground, physics-informed machine learning approaches have gained significant interest [7–16]. These approaches combine learned knowledge from data with the fundamental principles of imaging physics, offering a promising avenue for improved holographic image reconstruction. The methods can be classified into supervised [7–9], self-supervised [10–12], and untrained approaches [13–16]. Each has unique advantages and limitations, and the choice among them depends on specific requirements, data availability, and imaging conditions at hand.

Despite the promising performance of existing physics-informed machine learning approaches, their effectiveness heavily depends on the accuracy of the mathematical forward operator model. This reliance on the model's accuracy can be compromised by physical perturbations, such as imperfections or misalignments in the optical setup. As a result, the quality of reconstructed images may be degraded, leading to erroneous outcomes. In the context of digital holography, the propagation distance of the object wavefront is a crucial parameter. Obtaining precise knowledge of this parameter often necessitates calibration processes or autofocusing algorithms [21,22], which are both experimentally and computationally demanding.

Existing solutions to these challenges, while functional, present practical limitations. Lee *et al.*'s CycleGAN-inspired *supervised* approach [23], though enabling reconstruction beyond the

#550046 https://doi.org/10.1364/OE.550046
Journal © 2025 Received 25 Nov 2024; revised 19 Jan 2025; accepted 27 Jan 2025; published 5 Feb 2025

Optics EXPRESS

Table 1. Comparison of different holographic reconstruction methods

Methods	Traditional Iterative [4–6]	Physics-informed Machine Leaning [7–16]			Deep Learning [17–20]
		Supervised training [7–9]	Self-supervised training [10–12]	Untrained [13–16]	
Experimental data[a]	No need	Required	No need	No need	Required
Pre-training	No need	Required	Required	No need	Required
Inference efficiency	Low	High/Middle	High/Middle	Middle	High
Physical model	Required	Required	Required	Required	No need

[a]It refers to the actual complex-valued magnitude of the objects.

training data range, still necessitates extensive experimental measurements for prior knowledge acquisition. Similarly, the *self-supervised* GedankenNet [12], while eliminating the need for real training data, still requires pre-training using randomly generated artificial images. Moreover, the requirement of multiple input holograms introduces an inherent trade-off between reconstruction performance and system throughput, posing a potential limitation on its practical implementation when dealing with scarce measurements.

To completely eliminate the need for high-quality data from the target domain and the training process, researchers have explored *untrained* physics-informed machine learning methods [14–16]. These methods, exemplified by PhysenNet [16], leverage randomly initialized neural networks to parameterize the reconstructed object, minimizing a data consistency loss based on a predefined forward imaging model. Recent work [14] incorporates pre-trained diffusion models as generative priors for intensity reconstruction. These untrained neural network priors (UNNPs) effectively combine learned knowledge with the inherent imaging physics [24,25].

While UNNPs-based approaches reduce the need for extensive pre-training or high-quality target domain data, achieving successful reconstruction heavily depends on precise knowledge of the parameters in the forward operator [26–28]. Unfortunately, the issue of addressing the strong prior requirements in untrained reconstruction with UNNPs has not been thoroughly investigated yet. As a result, the performance of these methods remains limited and often requires significant effort, time, and expertise in terms of system alignments and calibration.

In this study, we present a transformative solution, called J-Net, for untrained physics-informed snapshot holographic imaging, delivering robust performance under deterministic perturbations regardless of propagation distance changes. In particular, J-Net enables simultaneous reconstruction of the complex-valued magnitude of the object and the propagation distance. Unlike existing solutions for supervised [7] and self-supervised [12] learning methods, J-Net does not involve any training process or dataset, relying solely on a single measurement for reconstruction. It consists of two distinct models responsible for predicting the complex-valued object magnitude and propagation distance, respectively. By estimating both quantities simultaneously, J-Net demonstrates superior robustness against disturbances in the distance parameter.

In the following sections, we present our experimental results validating the performance of J-Net in the simultaneous reconstruction of the complex-valued object magnitude and propagation distance. Compared with UNNPs-based methods requiring matched parameters, J-Net maintains consistent performance at arbitrary object-to-sensor distances. For thin objects with a uniformly distributed refractive index, such as biological cells, such amplitude and phase reconstruction can effectively provide three-dimensional information about the object. Furthermore, we evaluate J-Net on experimental samples without precise knowledge of the accurate distance, significantly reducing the labor-intensive requirements of system alignments and calibration. This work offers a solution to extend the performance and usability of untrained holographic reconstruction algorithms under physical perturbations in practical scenarios. We propose the integration of J-Net with established supervised and self-supervised training techniques as a means to advance

Optics EXPRESS

physics-informed machine learning in holographic imaging, alleviating concerns related to precise model alignment.

2. Method

2.1. Untrained neural network priors with joint optimization

We develop a framework called J-Net to combine the untrained neural network priors (UNNPs) [24] with a joint optimization strategy for addressing the physical model mismatch. As depicted in Fig. 1, J-Net consists of two distinct models: $M_{\boldsymbol{w}}(\cdot)$ and $M_{\boldsymbol{u}}(\cdot)$, with trainable parameters $\boldsymbol{w}$ and $\boldsymbol{u}$, respectively.

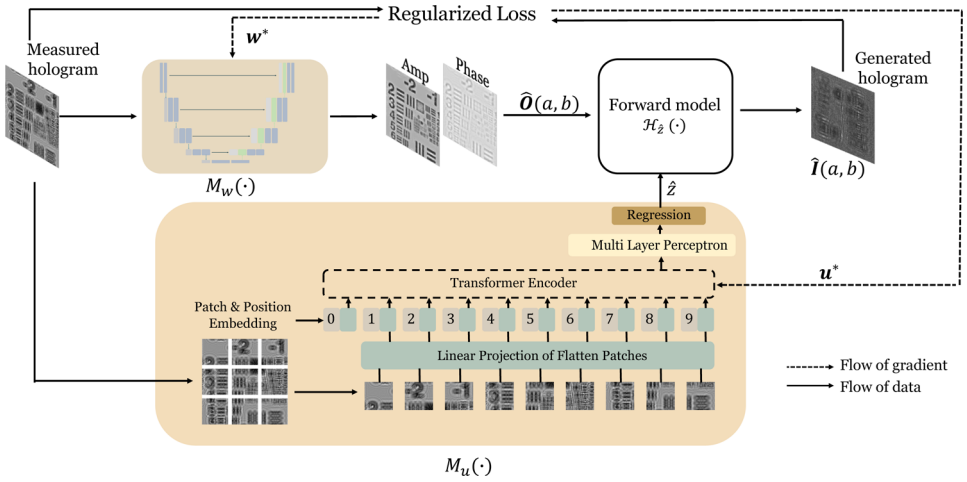

Fig. 1. Flowchart of the proposed joint optimization network (J-Net). It includes two jointly optimized networks: $M_{\boldsymbol{w}}$ and $M_{\boldsymbol{u}}$, responsible for predicting the complex-valued object magnitude and propagation distance, respectively. The trainable parameters $\boldsymbol{w}$ and $\boldsymbol{u}$ for each model are optimized through backpropagation using the regularized loss.

$M_{\boldsymbol{w}}(\cdot)$ is responsible for predicting the complex-valued object field $\hat{O}(a,b)$ from the measured hologram, i.e. $\hat{O}(a,b) = M_{\boldsymbol{w}}(I(a,b))$. It subsequently synthesizes a diffraction pattern $\hat{I}(a,b)$ through the forward imaging operator $\mathcal{H}_z(\cdot)$, which is defined using the angular spectrum method [1]:

$$\hat{I}(a,b) = \mathcal{H}_z(\hat{O}(a,b)) = \|\mathcal{F}^{-1}\left\{\mathcal{F}\left\{\hat{O}(a,b)\right\}\mathsf{H}_z(f_a,f_b)\right\}\|^2, \tag{1}$$

where $\mathcal{F}\{\cdot\}$ and $\mathcal{F}^{-1}\{\cdot\}$ are the Fourier and inverse Fourier transform, respectively. The transfer function is defined as $\mathsf{H}_z(f_a,f_b) = e^{j\frac{2\pi}{\lambda z}\sqrt{\frac{1}{\lambda^2}-f_a^2-f_b^2}}$ at each spatial frequency (f_a,f_b), which is parameterized by the propagation distance z, illumination wavelength λ, and sampling frequency determined by the pixel size. In traditional UNNPs-based methods, precise knowledge of the parameters in the forward operator is assumed. The reconstruction of the object's complex-valued magnitude involves minimizing the discrepancy between the synthetic hologram $\hat{I}(a,b)$ and the real observation $I(a,b)$:

$$O^*(a,b) = \underset{\hat{O}(a,b)}{\operatorname{argmin}}\|\hat{I}(a,b) - I(a,b)\|^2 = \underset{\hat{O}(a,b)}{\operatorname{argmin}}\|\mathcal{H}_z(\hat{O}(a,b)) - I(a,b)\|^2, \tag{2}$$

which is achieved by optimizing the parameters through the objective function:

$$\boldsymbol{w}^* = \underset{\boldsymbol{w}}{\operatorname{argmin}}\|\mathcal{H}_z(M_{\boldsymbol{w}}(I(a,b))) - I(a,b)\|^2. \tag{3}$$

Optics EXPRESS

It is clear that the objective function does not include the ground-truth object $O(a, b)$, which supports the use of a purely untrained mechanism based on a single-shot measurement. However, the assumption of precise forward modeling is frequently unmet due to inherent physical perturbations, leading to suboptimal performance of UNNPs-based methods. Among them, achieving reliable reconstruction typically requires precise alignment of the object's propagation distance, which can be a challenging and labor-intensive task. To address the strong prior requirements, our strategy is to jointly estimate the range of object, modifying the objective function to:

$$\boldsymbol{w}^*, z^* = \underset{\boldsymbol{w},z}{\operatorname{argmin}} \|\mathcal{H}_z(M_{\boldsymbol{w}}(I(a,b))) - I(a,b)\|^2. \tag{4}$$

By simultaneously predicting the object magnitude $O(a, b)$ and the disturbance factor z, J-Net effectively overcomes the challenges posed by propagation distance mismatch in UNNPs-based approaches for robust holographic imaging reconstruction. For the complex-valued object information, we explicitly represent it in terms of amplitude and phase, modeled as a dual-channel output of the network. This design inherently incorporates both amplitude and phase into the joint optimization process as unified objectives during training. Instead of representing the distance as a scalar, we propose to include another network for mapping it with the diffraction pattern, i.e., $\hat{z} = M_{\boldsymbol{u}}(I(a, b))$ with trainable parameters $\boldsymbol{u}$. Moreover, considering the edge-preserving smoothness prior of the object, we incorporate total variation regularization [29]. Overall, the two networks are optimized jointly by the objective function:

$$\begin{aligned} \boldsymbol{w}^*, \boldsymbol{u}^* &= \underset{\boldsymbol{w},\boldsymbol{u}}{\operatorname{argmin}} \|\mathcal{H}_{\hat{z}}(M_{\boldsymbol{w}}(I(a,b))) - I(a,b)\|^2 + \tau \|M_{\boldsymbol{w}}(I(a,b))\|_{TV}, \\ s.t., \quad \hat{z} &= M_{\boldsymbol{u}}(I(a,b)), \end{aligned} \tag{5}$$

where $\|\cdot\|_{TV}$ is the total variation regularization function [29] controlled by factor τ. In this way, the untrained mechanism is preserved, obviating the need for ground-truth object data while also relaxing the requirement for accurate parameter estimation in forward imaging model.

2.2. Network architecture and implementation

The overall structure of J-Net is demonstrated in Fig. 1 consisting of two distinct models: $M_{\boldsymbol{w}}(\cdot)$ and $M_{\boldsymbol{u}}(\cdot)$, responsible for predicting the complex-valued object magnitude and propagation distance, respectively.

$M_{\boldsymbol{w}}(\cdot)$ utilizes UNet [30] structure with dual-channel output, representing the amplitude and phase maps separately. The amplitude channel corresponds to the magnitude of the object field, while the phase channel represents the optical phase shift caused by the object. These two outputs are generated simultaneously, ensuring that the network effectively captures both the intensity and phase information of the object. By separating the amplitude and phase into individual real-valued maps, the network avoids the complexities associated with direct complex-value computation, allowing for more stable and interpretable training. As demonstrated in Fig. 2, the architecture comprises an "Hourglass" encoder and decoder structure with skip connections in the middle. The architecture incorporates four main types of operations to connect the input to the output. First, the down-convolution operation consists of a 3×3 convolution followed by batch normalization (BN) and LeakyReLU activation. This operation helps capture and extract relevant features from the input hologram. Second, max pooling is applied to downsample the feature maps, allowing for a larger receptive field and reduced spatial dimensions. Third, the up-convolution operation involves a 3×3 de-convolution followed by BN and LeakyReLU activation. This operation enables the upsampling and reconstruction of feature maps to match the input hologram resolution. Lastly, skip connections are employed to connect the corresponding feature maps from the encoder to the decoder. These connections facilitate the fusion of high-level and low-level features, allowing for better information flow and preserving fine details during

Optics EXPRESS

the reconstruction process. The final output is a tensor with two channels, each containing the reconstructed information of the corresponding map. To ensure valid ranges, a sigmoid layer is applied to limit the values within $[0, 1]$ for amplitude and $[0, \pi]$ for phase maps.

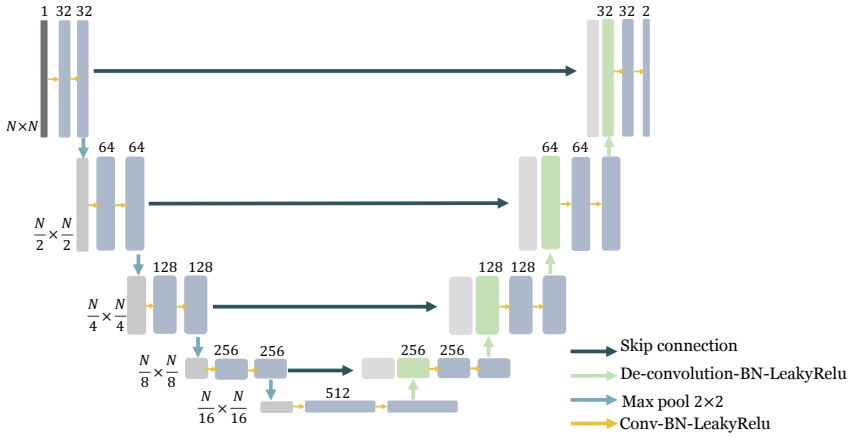

Fig. 2. Illustration of the architecture for $M_w(\cdot)$ in J-Net. The architecture comprises an "Hourglass" encoder and decoder structure with skip connections in the middle. The hologram is fed as input, and the model outputs the amplitude and phase maps.

$M_u(\cdot)$ is an additional network dedicated to jointly optimizing the physical perturbation factor in the parameterized model, specifically the distance variable z. Since each location of the hologram contains information about the propagation distance, we design $M_u(\cdot)$ based on the concept of self-attention from the Transformer architecture to capture dependencies between distant features compared to a pure CNN-based network [31,32].

As illustrated in Fig. 1, the hologram is divided into patches, which are linearly projected into an embedded dimension of $d_{model} = 48$. In Fig. 1, the division in the patching process is for demonstration purposes only, to offer a simplified illustration of the patching process while maintaining the overall concept. The actual number of patches depends on the model structure. In our study, we use a patch size of 8×8 for the input hologram with dimensions 512×512. This results in the number of patches, or the Transformer's sequence length, being $N = \frac{512 \times 512}{8^2} = 4096$. The embedded patches undergo processing by a transformer encoder as a sequence of vectors (tokens). Noteworthy, to regress the distance information, an additional position embedding for an extra learnable "distance token" is added, whose stage at the output from the transformer encoder represents the predicted distance. Figure 3 provides detailed insights into the internal structure of the transformer encoder, which consists of alternating Swin transformer blocks [33] and patch merging operations. The patch merging layer produces a hierarchical representation by concatenating the feature within the group of 2×2 neighboring with an initial patch size of 8×8. This arrangement enables better access to both local and global information within the input image compared to using a global window or the same patch size [34]. Swin transformer blocks [33] are applied for feature extraction, which consists of three major operations: normalization, multi-layer perceptions (MLP), and window-based multi-head self-attention (MSA).

As depicted in Fig. 3, the self-attention at each head is calculated using

$$\text{Attention}(Q, K, V) = \text{SoftMax}\left(\frac{QK^T}{\sqrt{d}}\right)V, \tag{6}$$

where $Q, K, V \in \mathbb{R}^{M \times d}$ are the query, key, and value matrices [31]. They are calculated by linearly projecting the input feature map with M patches using transformation matrices containing

Optics EXPRESS

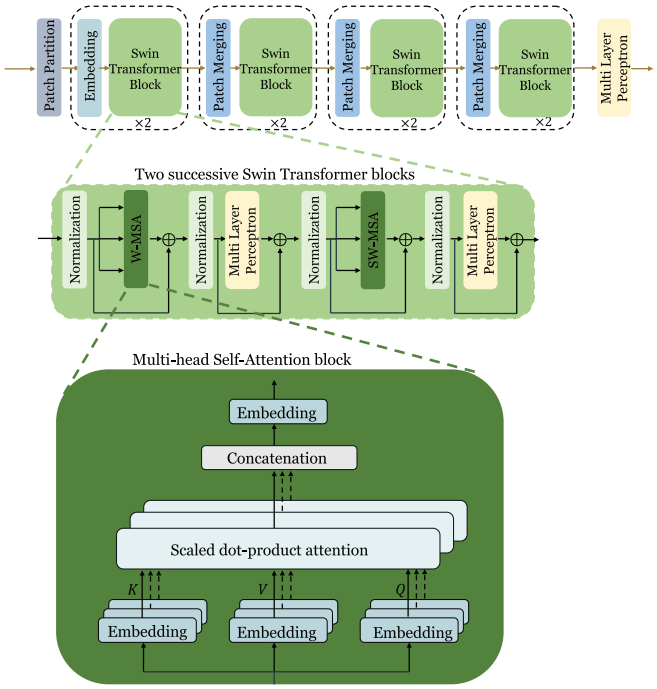

Fig. 3. Illustration of the transformer encoder architecture in $M_u(\cdot)$ of J-Net. It consists of multiple Swin Transformer blocks [33], arranged consecutively. Each block includes Multi-head Self-Attention (MSA) blocks that utilize regularly partitioned windows (W-MSA) and shifted windows (SW-MSA) for efficient attention-based feature extraction. The detailed mechanism within the MSA block [33] is demonstrated in the enlarged panel.

trainable parameters. The self-attention at each head is then concatenated and once again projected by the embedding layer to get the multi-head self-attention. Specifically, there are total 8 Swin transformer blocks [33] utilized in the transformer encoder, which are evenly separated by patch merging operations into 4 stages. The number of heads at each stage is set as $(2, 2, 4, 4)$, respectively. The trainable weights within this mechanism are jointly optimized to capture the internal feature relevance and merge information in subspaces, facilitating global dependence between the input and output.

It is worth noting that, MSA is calculated in a window-based scheme, which involves dividing the input feature into smaller, non-overlapping or overlapping windows and applying MSA among patches within each window separately. Moreover, to increase the cross-window connections for improved modeling power, the shifted window partitioning approach for calculating MSA is utilized that alternates between the regularly partitioned windows (W-MSA) and shifted windows (SW-MSA) in consecutive Swin transformer blocks with a window size of 16×16. Normalization is applied before window-based MSA blocks, and a residual connection is included. The multilayer perceptron is employed with a single hidden layer of $4\times$ expansion.

After processing through the transformer encoder, the multilayer perception is used to project the output to the scalar, and a bounded activation function (e.g., Sigmoid) is applied in the regression layer to prevent significant output deviations. This approach is reasonable since the actual propagation distance typically falls within a practical range.

The identical network structure is used for both simulated and experimental datasets to ensure consistency and demonstrate the model's generalizability. The network is implemented in Python, utilizing PyTorch and Numpy. Xavier normalization [35] is used for parameter initialization.

Optics EXPRESS

During each iteration, the loss is computed, and the trainable parameters $\boldsymbol{w}$ and $\boldsymbol{u}$ are updated through back-propagation using the Adam optimizer [36]. The Adam optimizer utilizes default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for exponential decay rates. To maintain consistent stability, the learning rate is kept fixed at 0.001. For most samples with holograms of a certain size 512×512, it takes approximately 1000 iterations to achieve satisfactory results. Under this situation, the implementation is accelerated by utilizing a single NVIDIA GeForce RTX 3070 GPU, resulting in a runtime of 1 minute.

3. Results

3.1. Reconstruction of complex amplitude and object distance

To demonstrate the simultaneous reconstruction of complex amplitude and object distance of J-Net, we conduct experiments from synthetic holograms where the actual propagation distances are available. It is important to note that the generated holograms serve as the only input to our network, while the provided ground truth distance is solely used for comparison purposes.

To generate synthetic complex object wavefields, we first convert the grayscale versions of these images, denoted as $\boldsymbol{x} \in [0, 1]$, to a resolution of 512×512 pixels. We then create three different cases: an amplitude-only object wavefield represented by $\boldsymbol{x}$, a phase-only object wavefield represented by $e^{j\pi\boldsymbol{x}}$, and an amplitude-phase object wavefield represented by $\boldsymbol{x} \odot e^{j\pi\boldsymbol{x}}$, where $\odot$ denotes elementwise multiplication. Using the forward propagation of holographic imaging introduced in Eq. (1), we generate simulated holograms and calculate the intensity of the resulting complex diffraction field for a given sample-to-sensor distance.

In our simulations, we consider the object that is illuminated by a coherent plane wave with a wavelength of 532 nm. The hologram is captured using a virtual camera with a resolution of 512×512 pixels, where each pixel corresponds to a physical size of 1.12 µm.

This aims to evaluate the feasibility of our model on various types of real-life targets. It is worth noting that the amplitude and phase information of the synthetic data is highly correlated that in accordance with the actual physics for some stained samples. Using the angular spectrum approach [1], we generate simulated holograms, and then calculate the intensity of the resulting complex diffraction field for a given sample-to-sensor distance. To visualize the convergence behavior of our algorithm, we plot the reconstruction results at three selected iterations in Fig. 4 with enlarged region-of-interest (ROI) for better visualization. As depicted in Fig. 4(a), 4(b), and 4(c), our algorithm gradually recovers the fine details and intricate structures of the object on all types of the object with the iteration increasing. They indicate J-Net's ability to reconstruct the complex-valued object wavefield without precise knowledge of the propagation distance. This ability is supported by its joint prediction of the sample-to-sensor distance, which is proofed by investigating the absolute error between the estimated distance $\hat{z}$ and the ground truth z using:

$$\text{Absolute error}\,(\%) = \frac{|\Delta z|}{z} = \frac{|z - \hat{z}|}{z}. \tag{7}$$

We plot the absolute error as a function of iteration for the three cases in Fig. 4(d)–4(f). J-Net exhibits convergence towards the true distance value in all cases, although the convergence speed varies. Notably, the absolute error curve for amplitude-phase objects demonstrates slower convergence and less stability. This finding highlights the inherent difficulties and challenges associated with holographic reconstruction for objects that exhibit complex transmittance properties.

3.2. Comparative study of robustness to perturbation

In this section, we investigate the extended robustness of UNNPs-based method towards distance perturbations with our joint optimization strategy. For a fair comparison, we select two untrained

Optics EXPRESS

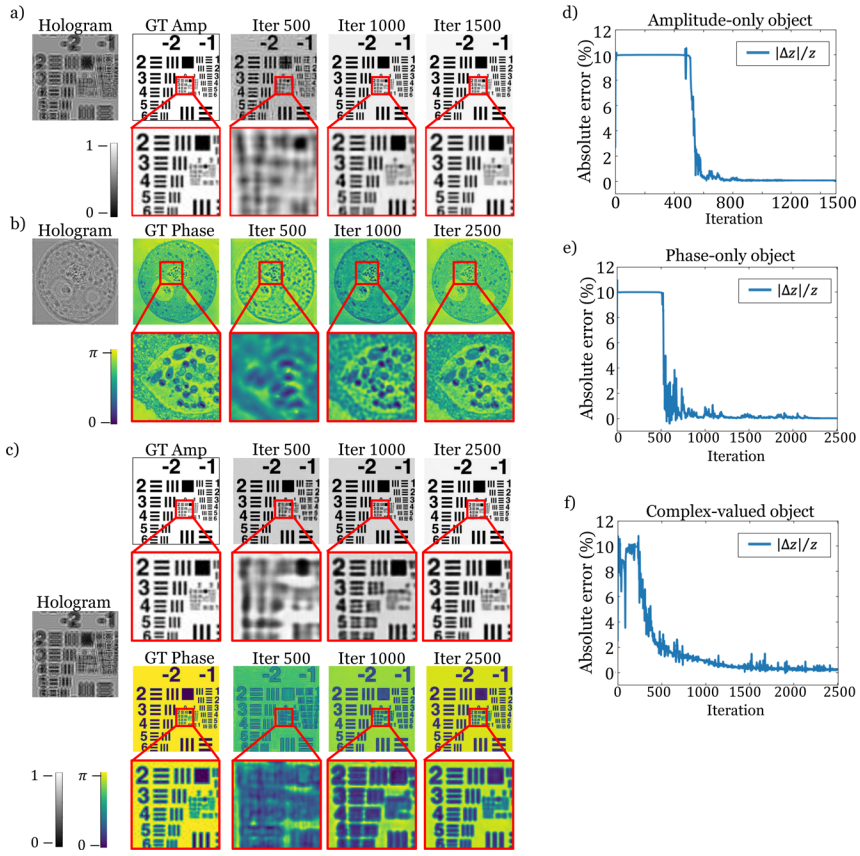

Fig. 4. Reconstruction process of different cases: (a) Amplitude-only object, (b) Phase-only object, and (c) Complex-valued object with corrected amplitude and phase maps. Holograms, GT maps, and reconstruction results at selected iterations are shown sequentially. Enlarged ROI regions are indicated by the red box in the whole FOV. (d-f) Absolute percentage difference between estimated and ground truth propagation distance during iterations for each case.

reconstruction methods, i.e., PhysenNet [16] and DeepDIH [15], which incorporate imaging physics. We implemented both methods from their official repositories, and for PhysenNet [16], we modified the network structure to provide double-channel outputs for amplitude and phase maps, respectively.

To conduct a comparative analysis, we set a propagation distance as a fixed parameter in the parameterized models required by PhysenNet [16] and DeepDIH [15]. This distance serves as the initial point in our J-Net model. We generate two sets of holograms under different conditions. In the first case, referred to as the *matched* case, we simulate holograms of objects using the actual base distance. This case represents an ideal scenario where the distance used for simulation perfectly matches the actual propagation distance. In the second case, referred to as the *mismatched* case, we intentionally introduce deviations from the base distance by randomly selecting distances for hologram generation. This case allows us to assess the performance of reconstruction methods when the distance used during simulation does not correspond exactly to the actual propagation distance. It is worth noting that we define a buffer range for the random selection of the J-Net's initial point to prevent significant distance deviations. This approach is

Optics EXPRESS

reasonable since the actual propagation distance generally remains within an approximate range from a practical standpoint.

For this comparison study, the base distance is set as $z = 1065$ μm with 20 % buffer area (i.e., ±200 μm). To quantitatively evaluate the performance, the structural similarity index measure (SSIM) is calculated to assess the similarity between the reconstructed amplitude and phase maps $\hat{\boldsymbol{x}}$ and their corresponding ground truth counterparts $\boldsymbol{x}$ for each method.

As demonstrated in Fig. 5, under the *matched* distance condition, both PhysenNet [16] and our method demonstrate comparable and similar performance when the physical model in the network matches the actual system, with outperforming DeepDIH [15]. However, under the *mismatched* distance, both PhysenNet [16] and DeepDIH [15] experience a significant decrease in the reconstruction quality. The arrows in Fig. 5(a) clearly indicate the performance gap between the methods. In contrast, our method maintains stable performance with SSIM values above 0.95 and 0.85 for amplitude and phase reconstruction, respectively, across the range of distances. For a more comprehensive demonstration, we examine a specific *mismatched* case with a 60 μm distance deviation ($z_{GT} = 1125$ μm), indicated by the red line in Fig. 5(a). Figure 5(b) illustrates the distance prediction of our model for $z_{GT} = 1065$ μm and $z_{GT} = 1125$ μm, respectively. As shown, J-Net gradually reconstructs the actual recording distance of the hologram for different cases, showcasing its ability to correct the physical imaging model in the network and successfully retrieve the actual propagation distance. The reconstruction results, plotted in Fig. 5(c), provide a detailed visual comparison. In PhysenNet [16], the SSIM values dropped from 0.96 to 0.66 for amplitude predictions and from 0.83 to 0.28 for phase predictions. DeepDIH [15] experiences a drop of 0.14 and 0.27 for amplitude and phase predictions, respectively. In terms of Mean Squared Error (MSE), PhysenNet and DeepDIH showed notable increases under distance mismatches, with MSE rising to 0.104 and 0.151 for amplitude predictions, and to 0.276 and 0.301 for phase predictions, respectively. These results underscore the robustness of J-Net to model mismatch compared to other physics-informed networks that employ parameterized imaging models. This is achieved through the automatic correction enabled by our joint optimization framework, which facilitates a smooth transition from the initial point to the actual propagation value. By dynamically adapting to the correct propagation distance, J-Net effectively mitigates the adverse effects of model mismatch in UNNPs-based method, leading to improved and consistent performance in hologram reconstruction.

To evaluate the robustness of J-Net with respect to variations in the initial z value, we conducted additional simulations where the initial z values were systematically varied around the ground truth ($z_{GT} \pm \Delta z$). For each z_{GT}, intensity and phase reconstructions were performed across 10 independent trials for each initialization value. Reconstruction quality was quantified using MSE and SSIM, and the results were analyzed to assess how reconstruction performance changes as the initial z deviates from z_{GT}. In Fig. 6, the trends and variability of the reconstruction quality are presented in the form of the mean and interquartile range (IQR) as a function of the relative deviation of the initial z value ($\pm\frac{\Delta z}{z}$). As shown in the figure, our method demonstrates consistent performance across varying levels of initial z deviation, as reflected by the stable mean values and narrow IQR for both MSE and SSIM metrics. While the IQR slightly broadens under extreme deviations, the overall robustness of the method remains evident, confirming its ability to effectively handle diverse initialization scenarios.

3.3. Performance on experimental samples

We validate the usability of J-Net in real experimental samples collected from a typical Gabor's inline holography system as shown in Fig. 7(a). A laser module (HNL100L, Thorlabs) is utilized to generate a single wavelength (632.8 nm) light source with a power output of 10 mW. To capture the holograms, we use a digital camera, specifically the NV-GE134GM-T model from MindVision, which features a CMOS sensor with a pixel size of 4.8 μm. The collimated and

Optics EXPRESS

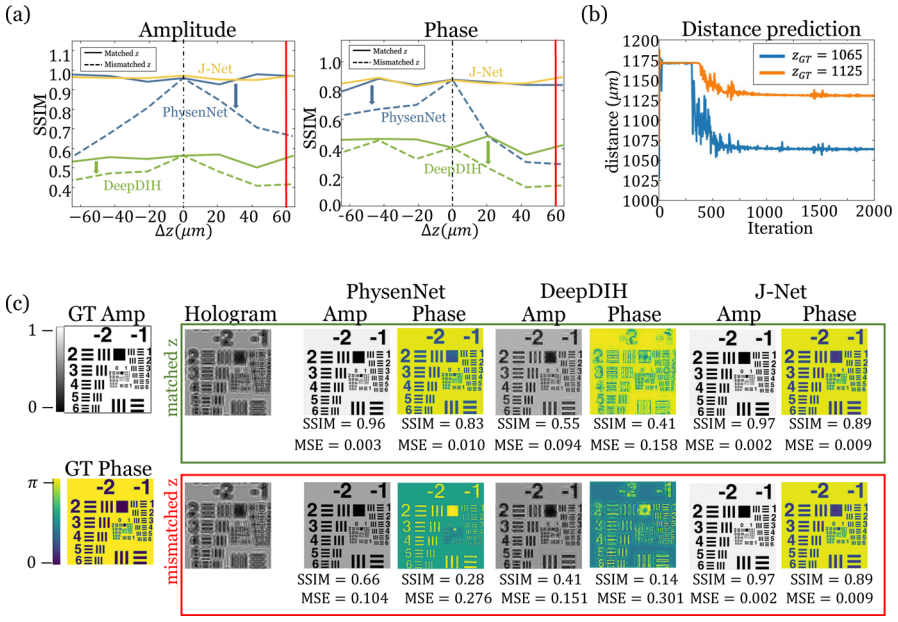

Fig. 5. Comparative study of robustness to perturbation. (a) Evaluation of reconstruction results on cases with *matched* and *mismatched* propagation distance z for PhysenNet [16], DeepDIH [15] and our method. Arrows indicate the performance gap between the two conditions for each method. A specific "mismatched" case (indicated by the red vertical line) with $\Delta z = 60$ μm ($z_{GT} = 1125$ μm) is highlighted for better inspection. (b) Distance prediction comparison of J-Net for both cases. (c) Comparison of reconstruction results between PhysenNet [16], DeepDIH [15], and J-Net.

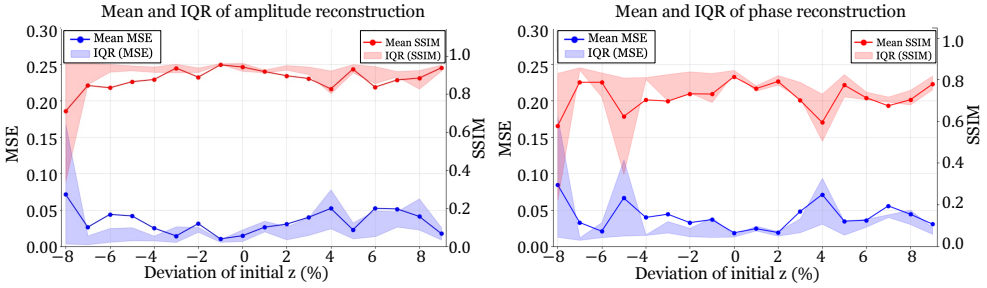

Fig. 6. Reconstruction quality as a function of the deviation of the initial z value ($\pm\frac{\Delta z}{z}$). The primary y-axis (left) shows the MSE, while the secondary y-axis (right) displays the SSIM. The solid lines represent the mean values across trials, while the shaded regions indicate the interquartile range (IQR).

expanded light passes through a beam expander (BE10M-A, Thorlabs) before interacting with the sample. For controlling the light intensity and preventing sensor overexposure, a neutral density filter (ND) is employed. The standard USAF intensity target (R1L1S7P, Thorlabs) is positioned along the light path, enabling the hologram to propagate onto the sensor plane. To validate the generalizability of our method across different system configurations, we utilized the biological samples, specifically a convallaria sample, collected using the optical system setup from Dr. Chen's research lab [37].

Optics EXPRESS

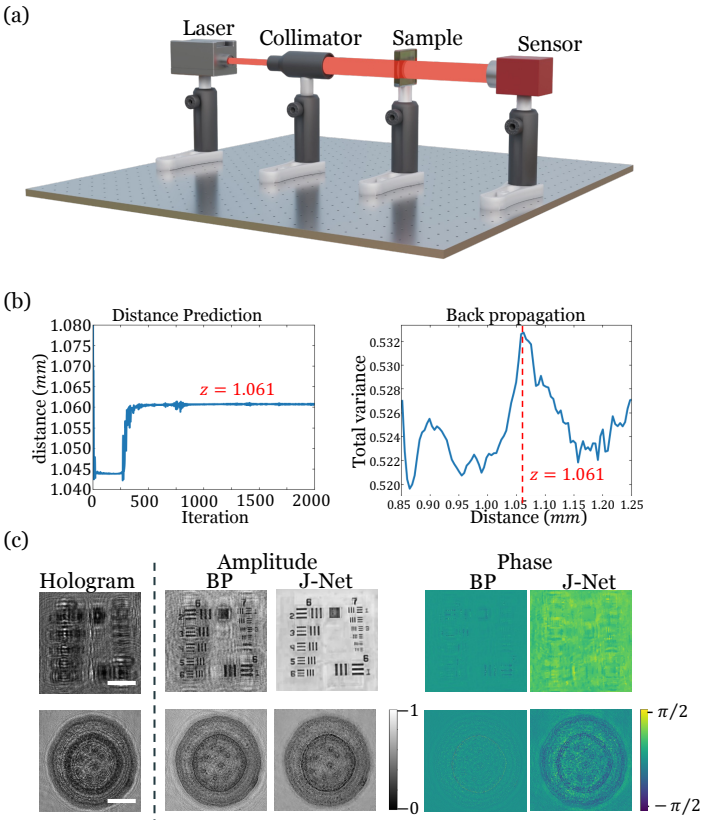

Fig. 7. Experimental verification. (a) Experimental setup of the optical system. (b) The distance prediction of J-Net on the experimental hologram (left) aligns with the value where the peak total variation (TV) occurs in the backpropagation reconstruction result (right). (c) Visualized reconstruction results on experimental samples, showing amplitude and phase maps for a USAF intensity object and biological samples. Scale bar is 0.7 mm.

As demonstrated in Fig. 7(b), the predicted distance converges to a value of $z = 1.061$ mm with iterative optimization. For validation purposes, we perform backpropagation along the axial axis and calculate the total variation (TV) value for each reconstructed result. The TV curve exhibits a distinct peak, indicating that our model accurately locates the distance of focus for the system. However, it should be noted that the BP method, which determines the focus distance based on the largest TV value, is only effective in simple scenarios, such as a resolution target with basic patterns. In cases involving complex objects or noisy holograms, the BP method faces significant limitations. These challenges stem from the inherent twin-image problem in inline holography, as BP does not explicitly separate real and twin images. This often results in significant distortions in reconstructed images, particularly when dealing with intricate diffraction patterns or high noise levels. Therefore, advanced algorithms, such as those proposed in our study, are essential to overcome these challenges, enabling precise focus distance determination and high-quality reconstructions, even in complex and noisy conditions. The jointly retrieved amplitude and phase maps are shown in Fig. 7(c), along with a comparison to the backpropagation result based on the predicted distance. With the intensity USAF target, our method successfully reconstructs the amplitude pattern, yielding a smoother background and sharper edges. The phase map demonstrates a roughly uniform pattern. On biological samples, in the backpropagation results, it is challenging to identify the focused pattern, particularly in the phase map. In contrast,

Optics EXPRESS

our approach successfully reconstructs high-quality amplitude and phase maps with detailed information. Besides, J-Net does not require precise knowledge of the propagation distance, which significantly reduces the labor-intensive process of system alignments and calibration in practical scenarios.

4. Conclusion

In conclusion, we have introduced J-Net, a novel approach for reliable and accurate untrained hologram reconstruction with automatic distance correction. By jointly optimizing the retrieval of complex-valued object magnitude and the actual propagation distance, J-Net achieves robust performance even with imprecise system calibrations. Comprehensive experiments demonstrate high-quality amplitude and phase reconstruction in both matched and mismatched distance scenarios, highlighting J-Net's resilience where other untrained methods falter. This ability to reconstruct complex-valued object fields from a single hologram, regardless of distance variations, promises to make digital inline holography a more accessible and versatile tool. Furthermore, the joint optimization framework of J-Net can be extended to address other deterministic perturbations, expanding the potential of untrained physics-informed methods. J-Net complements existing supervised [7] and self-supervised [12] learning approaches, establishing physics-informed machine learning as a powerful tool for practical holographic imaging.

Funding. Research Grants Council of Hong Kong (GRF 17200321, 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. Conflict of Interest The authors declare no competing interests.

Data availability. The code and data are available at [38].

References

1. J. W. Goodman, *Introduction to Fourier Optics* (W.H.Freeman & Company Ltd, 2017), 4th ed.
2. J. Huang, Y. Zhu, Y. Li, *et al.*, "Snapshot polarization-sensitive holography for detecting microplastics in turbid water," ACS Photonics **10**(12), 4483–4493 (2023).
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. Y. Rivenson, A. Stern, and B. Javidi, "Overview of compressive sensing techniques applied in holography," Appl. Opt. **52**(1), A423–A432 (2013).
5. W. Zhang, L. Cao, D. J. Brady, *et al.*, "Twin-image-free holography: a compressive sensing approach," Phys. Rev. Lett. **121**(9), 093902 (2018).
6. D. J. Brady, K. Choi, D. L. Marks, *et al.*, "Compressive holography," Opt. Express **17**(15), 13040–13049 (2009).
7. C. Lee, G. Song, H. Kim, *et al.*, "Deep learning based on parameterized physical forward model for adaptive holographic imaging with unpaired data," Nat. Mach. Intell. **5**(1), 35–45 (2023).
8. D. Yin, Z. Gu, Y. Zhang, *et al.*, "Digital holographic reconstruction based on deep learning framework with unpaired data," IEEE Photonics J. **12**(2), 1–12 (2020).
9. Y. Zhang, M. A. Noack, P. Vagovic, *et al.*, "PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets," Opt. Express **29**(13), 19593–19604 (2021).
10. X. Wu, Z. Wu, S. C. Shanmugavel, *et al.*, "Physics-informed neural network for phase imaging based on transport of intensity equation," Opt. Express **30**(24), 43398–43416 (2022).
11. J. Tang, J. Zhang, S. Zhang, *et al.*, "Phase aberration compensation via a self-supervised sparse constraint network in digital holographic microscopy," Opt. Lasers Eng. **168**, 107671 (2023).
12. L. Huang, H. Chen, T. Liu, *et al.*, "Self-supervised learning of hologram reconstruction using physics consistency," Nat. Mach. Intell. **5**(8), 895–907 (2023).
13. 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), 133701 (2023).
14. Y. Zhang, X. Liu, and E. Y. Lam, "Single-shot inline holography using a physics-aware diffusion model," Opt. Express **32**(6), 10444–10460 (2024).
15. 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).
16. F. Wang, Y. Bian, H. Wang, *et al.*, "Phase imaging with an untrained neural network," Light: Sci. Appl. **9**(1), 77 (2020).

Optics EXPRESS

17. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: A review," Opt. Express **29**(24), 40572–40593 (2021).
18. K. Wang, L. Song, C. Wang, *et al.*, "On the use of deep learning for phase recovery," Light: Sci. Appl. **13**(1), 4 (2024).
19. Y. Rivenson, Y. Zhang, H. Günaydın, *et al.*, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Sci. Appl. **7**(2), 17141 (2017).
20. 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).
21. Z. Ge, H. Wei, F. Xu, *et al.*, "Millisecond autofocusing microscopy using neuromorphic event sensing," Opt. Lasers Eng. **160**, 107247 (2023).
22. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica **5**(4), 337–344 (2018).
23. B. Sim, G. Oh, J. Kim, *et al.*, "Optimal transport driven CycleGAN for unsupervised learning in inverse problems," SIAM J. Imaging Sci. **13**(4), 2281–2306 (2020).
24. A. Qayyum, I. Ilahi, F. Shamshad, *et al.*, "Untrained neural network priors for inverse imaging problems: A survey," IEEE Trans. Pattern Anal. Mach. Intell. **45**, 1–20 (2022).
25. K. Wang and E. Y. Lam, "Deep learning phase recovery: data-driven, physics-driven, or a combination of both?" Adv. Photonics Nexus **3**(05), 056006 (2024).
26. Z. Burns and Z. Liu, "Untrained, physics-informed neural networks for structured illumination microscopy," Opt. Express **31**(5), 8714–8724 (2023).
27. 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).
28. I. Kang, M. De Cea, J. Xue, *et al.*, "Simultaneous spectral recovery and cmos micro-led holography with an untrained deep neural network," Optica **9**(10), 1149–1155 (2022).
29. A. Chambolle and P.-L. Lions, "Image recovery via total variation minimization and related problems," Numerische Mathematik **76**(2), 167–188 (1997).
30. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of Medical Image Computing and Computer-assisted Intervention (MICCAI)*, pp. 234–241 (2015).
31. A. Vaswani, N. Shazeer, N. Parmar, *et al.*, "Attention is all you need," Advances in Neural Information Processing Systems **30** (2017).
32. S. Cuenat, L. Andréoli, A. N. André, *et al.*, "Fast autofocusing using tiny transformer networks for digital holographic microscopy," Opt. Express **30**(14), 24730–24746 (2022).
33. Z. Liu, Y. Lin, Y. Cao, *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF International Conference on Computer Vision* pp. 10012–10022 (2021).
34. A. Dosovitskiy, L. Beyer, A. Kolesnikov, *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv (2020).
35. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics* pp. 249–256 (2010).
36. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *Proceedings of the International Conference on Learning Representations.* (2015).
37. N. Chen, C. Wang, and W. Heidrich, "∂**H**: Differentiable holography," Laser Photonics Reviews **17** (2023).
38. Y. Zhang, "Robust holographic imaging for real-world applications with joint optimization," (2025). https://github.com/yp000925/J-Net.