Engineering Applications of Artificial Intelligence 163 (2026) 112970

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Simple and practical single-shot digital holography based on unsupervised diffusion model

Seonghwan Park [a], Jaewoo Park [b], Youhyun Kim [a], Inkyu Moon [a,b,*], Bahram Javidi [c]

[a] Department of Robotics & Mechatronics Engineering, DGIST, Daegu, 42988, South Korea
[b] Department of Artificial Intelligence, DGIST, Daegu, 42988, South Korea
[c] Department of Electrical and Computer Engineering, University of Connecticut, Storrs, CT, 06269, USA

ARTICLE INFO

Keywords:
Digital holography
Phase image reconstruction
Unsupervised diffusion model
Deep learning
Computational imaging

ABSTRACT

Single-shot digital holography in Gabor mode offers cost-effective quantitative phase imaging but suffers from the fundamental twin image problem, where real and conjugate images are inherently superimposed, severely limiting phase reconstruction accuracy. Traditional iterative phase retrieval methods require computationally expensive multiple propagations, while off-axis holography demands complex optical setups with precise alignment. We present the first unsupervised diffusion model for automated phase image reconstruction from single-shot in-line holograms, eliminating both twin image artifacts and the need for expensive off-axis configurations. Our framework integrates cycle-consistency and denoising modules to enable training on unpaired hologram-phase image datasets, learning the mapping between low-cost in-line measurements and high-quality phase distributions without requiring labeled data pairs. Comprehensive evaluation on diverse biological specimens demonstrates that our approach significantly outperforms conventional unsupervised methods, achieving superior Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) values for both red blood cells and cancer cells. Critically, the model maintains exceptional performance even with limited training data, consistently outperforming supervised learning approaches under data-constrained conditions. The framework exhibits remarkable generalization capabilities, successfully reconstructing phase images from holograms captured at different propagation distances and processing various cancer cell types not included in training data. This computational breakthrough enables accurate, scalable, and hardware-efficient quantitative phase imaging, democratizing access to high-quality phase microscopy for resource-constrained environments while maintaining reconstruction fidelity comparable to complex off-axis systems.

1. Introduction

The single-shot holography technique was first proposed by Gabor in 1948 (Gabor, 1948). Since it uses fewer optical components and is relatively easy to install, single-shot holography with Gabor configuration without a separate reference beam has been applied to various fields, including biological applications (Moon and Javidi, 2008; Xu et al., 2001) and optical authentication (Chen, 2019). In Gabor mode, as originally identified by Gabor, the real and conjugate (twin) images are inherently superimposed within the recorded hologram—a fundamental characteristic that persists in both analog and digital implementations. To decouple these overlapping components from a single acquisition, iterative phase retrieval methods have been introduced, leveraging measurement diversity and prior knowledge of the sample (Koren et al., 1993; Latychevskaia, 2019; Latychevskaia and Fink, 2007; Nakamura et al., 2007).

However, to reconstruct the actual phase value, light must be propagated back and forth several times, so the time consumption is large. Off-axis holography introduces a slight angular offset between the object and reference waves, which enables the separation and removal of the conjugate image in the frequency domain, thereby allowing accurate extraction of quantitative phase information from the sample (Cuche et al., 1999; Javidi et al., 2021; Kemper and von Bally, 2008; Rappaz et al., 2005). Moreover, off-axis holography enables the acquisition of label-free phase images of living cells suspended in a liquid medium, while minimizing radiation exposure. The phase information can be retrieved from a single hologram without the need for mechanical scanning, thereby facilitating three-dimensional (3D) reconstruction of

* Corresponding author. Department of Robotics & Mechatronics Engineering, DGIST, Daegu, 42988, South Korea.
E-mail address: inkyu.moon@dgist.ac.kr (I. Moon).

https://doi.org/10.1016/j.engappai.2025.112970
Received 21 April 2025; Received in revised form 30 August 2025; Accepted 23 October 2025
Available online 29 October 2025

the specimen. Digital holography is used to study red blood cells (Bhabhor et al., 2024; Jaferzadeh and Moon, 2016; Jaferzadeh et al., 2018; Rappaz et al., 2008; Yi et al., 2015), white blood cells (Ugele et al., 2018), cancer cells (Lam et al., 2018; Pirone et al., 2022; Roitshtain et al., 2017), and cardiomyocytes (Ahamadzadeh et al., 2022; Shaked et al., 2010).

Despite the information from holography being very useful, many biological environments still present certain challenges in the wide applicability of off-axis holography. To obtain an image with perfect phase value through numerical reconstruction, off-axis holography must be equipped with a precise optical environment. Prior to recording holograms of biological samples, the optical paths of the reference and object waves must be carefully aligned. Any modifications introduced in the object arm must be correspondingly adjusted in the reference arm to maintain proper interference conditions. The other challenging steps in holographic image reconstruction are phase unwrapping (Herráez et al., 2002; Pritt and Shipman, 1994; Quiroga and Bernabeu, 1994), and digital correction of the phase aberrations (Colomb et al., 2006). While phase unwrapping algorithms can be applied to reconstruct continuous phase maps from wrapped phase images, these additional processing steps introduce computational complexity and potential sources of error, particularly in low signal-to-noise ratio conditions or when dealing with complex biological structures. Without proper application of these correction algorithms, it becomes challenging to extract accurate quantitative phase information from the sample.

1.1. Previous work and limitations

Recently, various state-of-the-art studies have produced meaningful results by applying deep learning to digital holography (Barbastathis et al., 2019; Moon, 2022). Several studies have focused on cellular analysis applications, including spatio-temporal cellular identification and pathological diagnosis (O'Connor et al., 2020; O'Connor et al., 2021), red blood cell segmentation and classification (Kim et al., 2022), and multi-depth phase-only hologram generation using convolutional neural networks (Yan et al., 2024). Furthermore, extensive research has been conducted on deep learning approaches for holographic phase reconstruction, with numerous studies demonstrating the potential of neural networks in recovering phase information from holographic data. Dardikman-Yoffe et al. (2020) developed PhUn-Net, a multi-layer encoder-decoder residual convolutional neural network specifically for phase unwrapping in biological cells, demonstrating robustness across various cell types. Jaferzadeh and Fevens (2022) proposed HoloPhaseNet using conditional generative adversarial networks for automated hologram reconstruction, achieving high accuracy in phase recovery tasks. Park et al., 2023, 2024 developed supervised models for both automated quantitative phase reconstruction and super-resolution enhancement in digital holography, successfully achieving high reconstruction quality. Rivenson et al. (2019) introduced a pioneering supervised deep learning framework for phase recovery and holographic reconstruction using neural networks, demonstrating remarkable performance in eliminating twin-image artifacts and reconstructing phase information from single holograms. These supervised approaches have shown impressive capabilities in various holographic imaging applications.

However, the majority of these deep learning approaches for holographic phase reconstruction have been developed within the supervised learning framework, each with inherent limitations that constrain their practical deployment. PhUn-Net required extensive paired datasets of wrapped and unwrapped phase images for training, while HoloPhaseNet necessitated substantial labeled training data and remained limited to specific optical configurations. The supervised models by Park et al. suffered from poor generalization when training data was limited and required computationally expensive data collection processes. Similarly, Rivenson et al.'s approach, despite its remarkable performance, required extensive paired training datasets, making it challenging to adapt to new sample types or imaging conditions without retraining.

The reliance on supervised learning creates several fundamental challenges for practical implementation. First, acquiring large-scale paired datasets is labor-intensive and costly, particularly for biological applications where ground truth phase images must be obtained through complex off-axis holographic setups. Second, supervised models typically exhibit poor generalization to sample types not included in the training data, limiting their versatility across diverse biological specimens. Third, the requirement for extensive labeled data becomes prohibitive when dealing with limited data scenarios, which are common in specialized biological research.

While some unsupervised approaches have been proposed, they face significant limitations. Li et al. (2020) demonstrated unsupervised auto-encoder methods for single-shot digital in-line holography reconstruction, but their iterative approach requires multiple back-and-forth light propagations, resulting in substantial computational time for generating phase images from single holograms. Similarly, Manisha et al. (2023) showed twin image removal using random illumination-based recording with unsupervised auto-encoders, but this method also relies on iterative processing that significantly slows down single-shot phase reconstruction. Yin et al. (2020) developed a CycleGAN-based framework for digital holographic reconstruction using unpaired data, incorporating cycle consistency loss and generative adversarial networks. While their approach demonstrated robustness against displacement aberrations and defocusing effects, CycleGAN-based methods typically suffer from training instability and mode collapse issues. Zhang et al. (2021) introduced PhaseGAN, a generative adversarial network-based approach for unpaired phase retrieval that incorporates physics of image formation. However, conventional unsupervised techniques like CycleGAN (Zhu et al., 2017), UNIT (Liu et al., 2017), and even PhaseGAN often struggle with the complex biological holographic data characteristics, particularly when dealing with diverse cellular morphologies simultaneously, highlighting the need for more robust unsupervised learning frameworks specifically designed for holographic phase reconstruction challenges.

1.2. Contribution

In this study, we present a groundbreaking computational framework that, for the first time in the field of digital holography, employs an unsupervised diffusion model for automated phase image reconstruction from single-shot in-line holograms recorded in cost-effective Gabor mode. Our approach addresses a fundamental challenge in holographic imaging: achieving high-quality quantitative phase reconstruction from low-cost, simple optical setups without requiring paired training data. The trained model operates as a true single-shot system, enabling rapid phase image acquisition by simply inputting an in-line hologram, which dramatically improves computational efficiency compared to traditional iterative phase retrieval methods that require multiple back-and-forth light propagations. This breakthrough eliminates the need for complex and costly off-axis holographic systems while maintaining reconstruction fidelity, making quantitative phase imaging accessible for resource-constrained environments and real-time applications.

Our unsupervised diffusion-based approach demonstrates exceptional robustness and generalization capabilities that surpass existing methods. The framework exhibits remarkable adaptability across varying optical conditions, successfully reconstructing phase images from holograms captured at different propagation distances beyond the training range, addressing practical scenarios where precise distance control is challenging. Furthermore, the model shows superior generalization to diverse cellular morphologies, effectively processing various cancer cell types that were not included in the training dataset, alongside the original red blood cells and ovarian cancer cells used for training. Through comprehensive experimental validation, we demonstrate that our method significantly outperforms conventional

unsupervised techniques such as CycleGAN and UNIT while maintaining exceptional performance even with limited training data. This combination of speed, robustness, and generalization capability represents a paradigm shift in holographic phase reconstruction, potentially democratizing access to high-quality quantitative phase microscopy for biological research and clinical applications.

The contributions of the proposed scheme can be summarized as follows:

1. This study presents the first computational framework in digital holography that pioneers the use of an unsupervised diffusion model for automated phase image reconstruction from single-shot Gabor holograms, marking a significant advancement in data-driven phase reconstruction for digital holography.
2. Our approach achieves hardware–software co-optimization, enabling cost-effective quantitative phase imaging by transforming simple in-line holographic measurements into high-quality phase images. This allows off-axis-equivalent reconstruction performance from a significantly simpler and more compact optical system, thereby reducing complexity and cost while preserving quantitative fidelity.
3. The framework demonstrates superior reconstruction quality and robustness compared to conventional unsupervised methods such as CycleGAN and UNIT, while achieving exceptional performance even with limited training data, consistently outperforming supervised learning approaches under data-constrained conditions.
4. The trained model operates as a true single-shot system with rapid inference capabilities, dramatically outperforming traditional iterative phase retrieval methods in computational efficiency while maintaining high reconstruction fidelity for real-time applications.
5. The method exhibits exceptional generalization capabilities, successfully reconstructing phase images from holograms captured at different propagation distances and processing various cancer cell types not included in the training dataset, demonstrating remarkable adaptability to new optical conditions and cellular morphologies.

2. Method

2.1. Sample preparation

Peripheral blood samples were collected from three healthy donors via venipuncture and immediately mixed with HEPA buffer. The diluted blood was then applied onto imaging slides. All procedures were conducted in compliance with institutional guidelines and regulations (DGIST-180713-BR-012-01) and approved by the DGIST bio-safety committee and IRB in Korea. Experiments were completed within a few hours of collection, after which the samples were discarded correctly. For cancer cell imaging, the ovarian cancer cell line (SK-OV-3, HTB-77, ATCC) with epithelial morphology was used. Cells were maintained in 35 mm dishes with low walls and a polymer coverslip base (Ibidi 80136). They were cultured under standard conditions at 37 °C with 5 % CO_2 in a humidified incubator. The growth medium consisted of BI RPMI 1640 (ATCC 30–2001), supplemented with 10 % fetal bovine serum (ATCC 30–2020).

2.2. In-line (gabor) and off-axis digital holographic imaging

Holographic imaging was carried out using a conventional digital holographic microscopy (DHM) setup, a configuration that is widely recognized and validated in prior studies (Cuche et al., 1999). The optical system was equipped with two Leica microscope objectives: a 20 × objective with numerical aperture 0.55 and a 40 × objective with numerical aperture 0.80, providing pixel sizes of 0.2849 μm and 0.1441 μm, respectively. Coherent illumination was supplied by a red diode laser operating at 666 nm wavelength. Image detection was accomplished using a CMOS camera with a sensor resolution of 1900 × 1200 pixels and individual pixel dimensions of 5.86μm × 5.86 μm. The hologram size was cropped to 1024 × 1024 pixels to optimize computational efficiency. All acquisitions were performed in a controlled laboratory environment with temperature stabilization to minimize thermal drift and mechanical vibrations.

Fig. 1 illustrates the general layout of in-line and off-axis digital holographic microscopy and demonstrates the fundamental differences between these two approaches. In our optical setup, the microscope objective (MO) is positioned immediately after the specimen to collect and magnify the diffracted object wave. For in-line hologram acquisition, only the object wave was recorded through its interference with the unscattered illumination. The off-axis holography realized using a Mach–Zehnder interferometer, introduces a slight angular separation between the object and reference waves. This angular separation enables spatial frequency domain separation of the real image, twin image, and zero-order terms (Cuche et al., 1999). While off-axis holography provides twin image-free reconstruction, it requires precise optical alignment, beam path matching, and additional optical components including beam splitters and mirrors, resulting in significantly higher optical complexity and cost compared to the simplified in-line approach. To facilitate experiments, our system was designed to allow rapid switching between in-line and off-axis modes simply by operating the beam shutter, enabling efficient data collection under identical imaging parameters. It should be noted that while a basic digital holographic microscopy (DHM) without an MO offers simplicity, its transverse resolution is fundamentally limited by the sensor pixel size and recording geometry, typically several micrometers. By contrast, incorporating an MO enables the reconstructed phase images to achieve a resolution approaching the diffraction limit of the imaging system, allowing transverse resolution below 1 μm. As established in prior work (Cuche et al., 1999), MO-based DHM can be regarded as a form of image holography, since the hologram is recorded from the magnified intermediate image produced by the MO rather than directly from the object.

Phase and amplitude images were reconstructed from off-axis holograms using established numerical propagation methods following the framework described by Cuche et al. (1999). Detailed mathematical formulations and implementation parameters of the reconstruction procedure are provided in the Supplementary Material S1. The reconstruction yields both amplitude images and phase images. Since the obtained phase values are typically wrapped within the [-π, π] interval, quality-guided phase unwrapping algorithms (Herráez et al., 2002) were applied when phase variations exceeded 2π to reconstruct accurate quantitative phase maps. The resulting phase images serve as reliable reference data, validated through established reconstruction algorithms and system specifications (Cuche et al., 1999).

For transparent biological specimens such as living cells, phase information typically provides superior contrast and quantitative data compared to amplitude, as cellular structures exhibit minimal absorption but significant refractive index variations. Phase information provides quantitative optical path difference measurements directly related to sample thickness and refractive index variations, enabling calculation of important cellular parameters including dry mass, projected area, and cell volume (Roitshtain et al., 2017). Therefore, in this study, we focused only on phase reconstruction, which is most important for quantitative biological imaging applications, in model training.

2.3. Data generation

For model training, we collected two types of optical data: in-line holograms recorded in Gabor mode and high-quality phase images obtained from off-axis digital holographic microscopy. This training strategy enables our unsupervised diffusion model to learn the mapping relationship between cost-effective in-line holographic measurements and accurate quantitative phase information without requiring paired datasets. The off-axis phase images serve as reliable reference targets during the training process, providing ground truth phase information

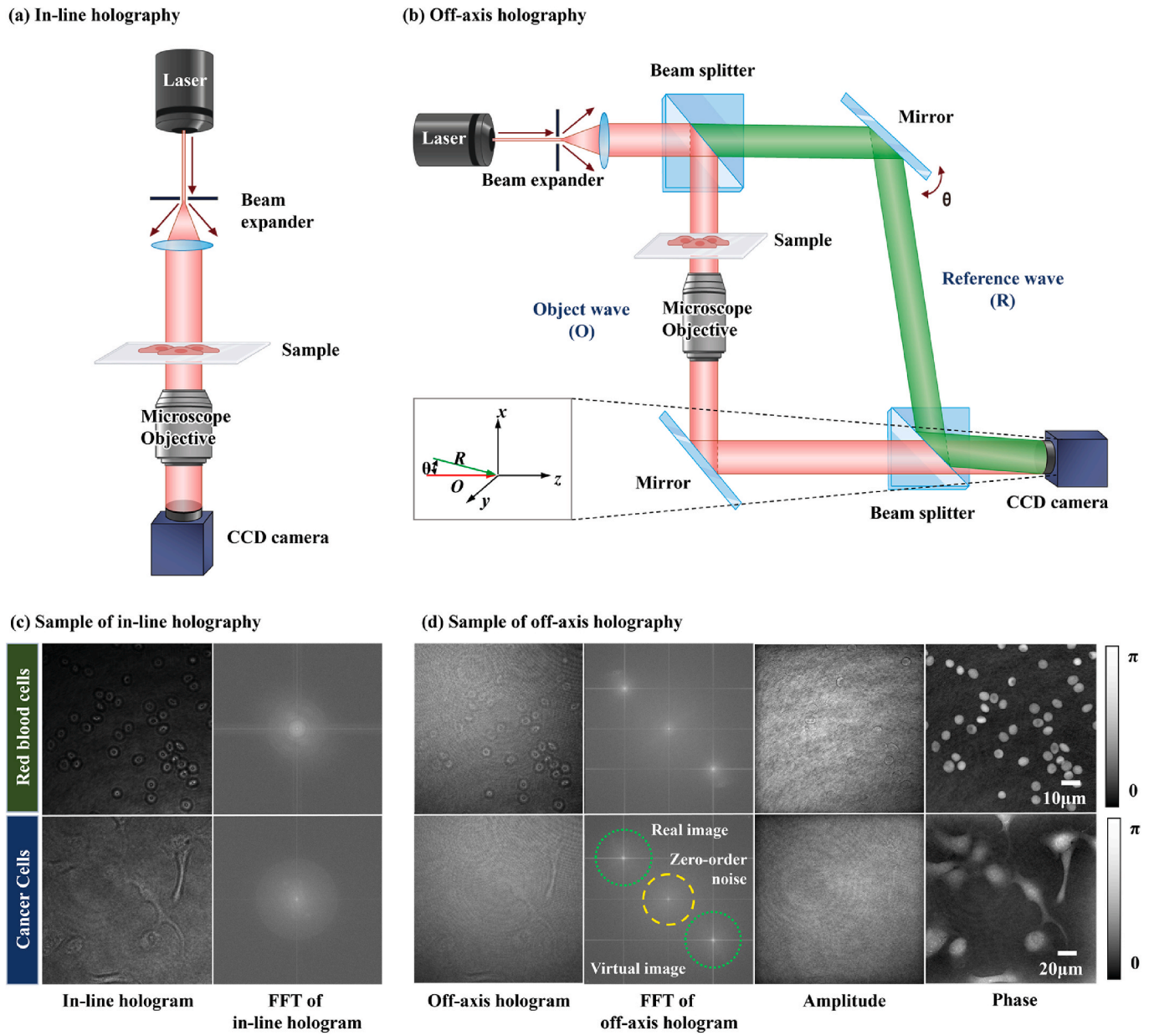


Fig. 1. General layout of (a) in-line and (b) off-axis digital holographic microscopy. (c) Sample images from in-line holography showing recorded holograms and their corresponding FFT spectra for red blood cells and cancer cells. (d) Sample images from off-axis holography showing recorded holograms, FFT spectra, and reconstructed amplitude and phase images. The phase distributions include quantitative intensity bars indicating the range of reconstructed phase values in radians. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

with high accuracy validated by established reconstruction algorithms and system specifications. Importantly, once the model is trained, it operates as a true single-shot system requiring only a single in-line hologram input to generate the corresponding phase image, eliminating the need for off-axis measurements during inference.

In-line and off-axis holograms of red blood cells and cancer cell lines were acquired using digital holographic microscopy (DHM) (Fig. 2). Off-axis holograms were captured via a Mach–Zehnder interferometric DHM setup and reconstructed into phase images through numerical propagation based on Fresnel diffraction. In-line holograms were obtained by blocking the reference beam in the same system, thus recording only the object wave. The holograms were recorded at 1024 × 1024 pixels, while phase images reconstructed from off-axis holograms had a resolution of 900 × 900 pixels due to the circular filtering process that isolates the useful diffraction information and eliminates noisy edge regions. The reconstructed images covered approximately 129.68 μm × 129.68 μm for red blood cells imaged at 40 × magnification and 256.38 μm × 256.38 μm for cancer cells imaged at 20 × magnification. To ensure consistent region of interest for model training, the in-line holograms were cropped to 900 × 900 pixels to match the phase image dimensions. We recorded 900 holograms of red blood cell samples, and 750 holograms of cancer cells. Using data augmentation, such as rotation and flip, we created datasets containing 3600 images of red blood cell and 3000 images of cancer cells, respectively. The test dataset consisted of images excluded from the training process and was used solely to evaluate the model's performance. The test dataset was used by selecting 200 images from each sample's dataset, with the rest all being used as the training dataset.

2.4. Unsupervised diffusion model for phase reconstruction

This study presents an unsupervised diffusion model that directly

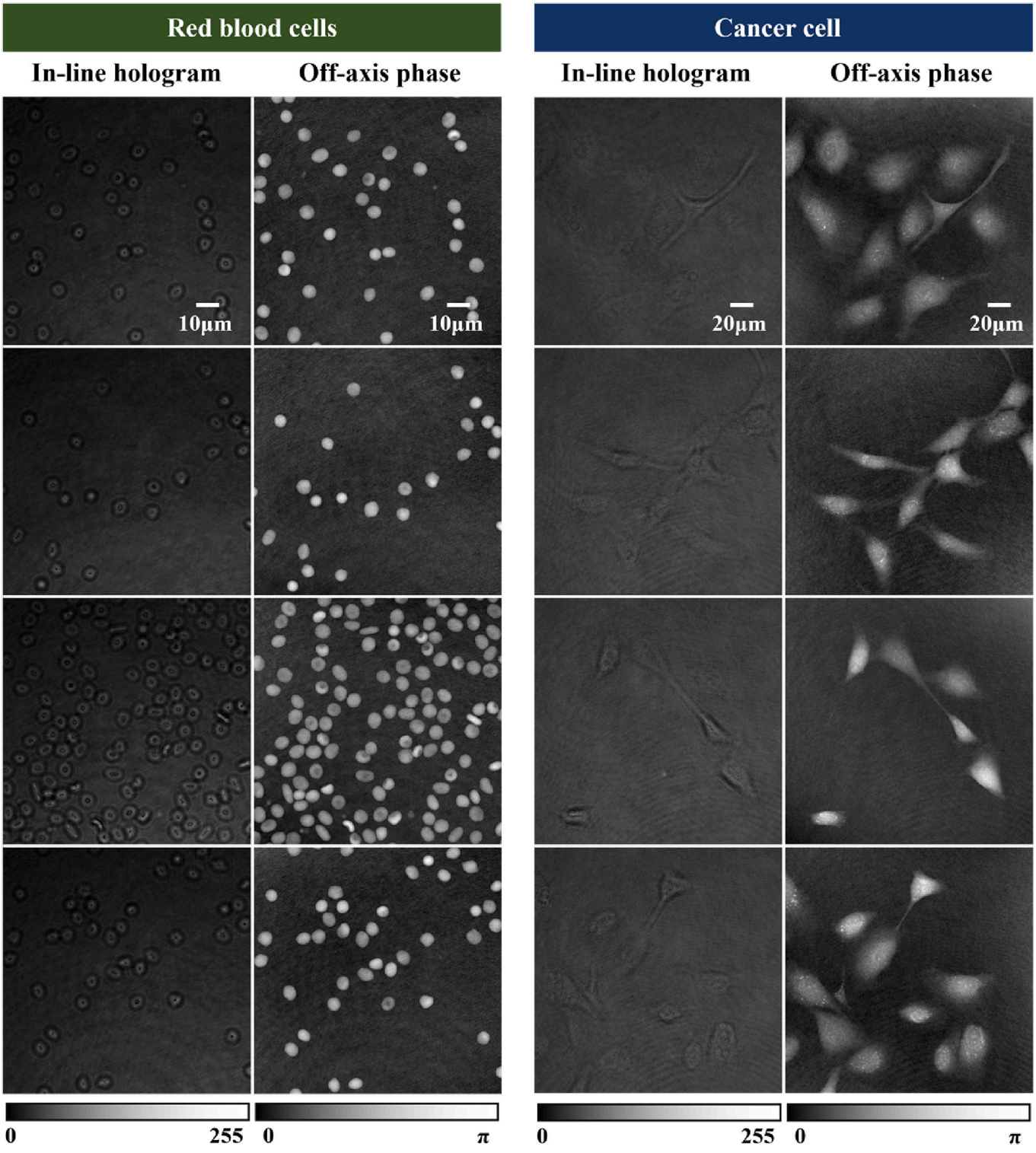


Fig. 2. Representative images of red blood cells and cancer cells used for model training. Each in-line hologram and corresponding phase image have a resolution of 900 × 900 pixels, covering an area of approximately 129.68 μm × 129.68 μm for red blood cells and 256.38 μm × 256.38 μm for cancer cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

reconstructs quantitative phase images from in-line holograms, solving the twin image problem without requiring complex off-axis optical configurations. The model utilizes two types of optical data during training: in-line holograms and clean phase images obtained from off-axis holography. Based on Özbey et al. (2023), this approach enables unsupervised learning where datasets do not require direct correspondence between specific hologram-phase image pairs. The model learns the general relationship between holographic patterns and phase distributions using the entire dataset. The computational framework consists of two interconnected networks. First, a cycle-consistency module learns bidirectional translation between hologram and phase domains using cycle-consistency loss. Second, a diffusion module performs iterative noise removal to generate high-quality phase images or holograms by progressively refining the reconstruction over multiple steps.

The model is trained for phase reconstruction, but to implement unsupervised learning, both hologram-to-phase and phase-to-hologram transformations are learned during training. However, the primary function of the trained model is phase reconstruction from real optical holograms. The phase-to-hologram transformation serves merely as a computational mechanism that generates hologram-like patterns consistent with the input holographic data to enable unpaired training. Once training is completed, the model operates as a true single-shot system that generates corresponding phase images from single in-line holograms. The trained diffusion module, learned for generating phase images from in-line holograms, reconstructs phase images from actual captured in-line holograms.

This approach addresses several practical optical challenges by eliminating the need for precise beam alignment required in off-axis systems and reducing optical complexity and cost by removing additional mirrors and beam splitters. The innovation lies in integrating computational techniques to solve a specific optical problem that has limited the practical deployment of in-line holography, while maintaining quantitative phase accuracy comparable to established optical methods.

2.4.1. Model architecture

2.4.1.1. Cycle-consistency module. The cycle-consistency module addresses a fundamental challenge in unsupervised learning for holographic reconstruction: how to learn meaningful mappings between in-line holograms and phase images without paired training data. This module establishes bidirectional translation between the hologram domain and phase domain by enforcing a mathematical constraint known as cycle-consistency.

Fig. 3 shows the process of the cycle-consistency module. The original phase images are obtained from off-axis holography, while the original in-line holograms are acquired from in-line holography. The module operates on the principle that if a hologram is transformed into the phase domain and then back into the hologram domain, the result should closely match the original hologram. Similarly, if a phase image is transformed into the hologram domain and back to the phase domain, it should return to its original form. This bidirectional constraint prevents the model from learning arbitrary or physically meaningless mappings between the two domains. The cycle-consistency module comprises two sets of generator–discriminator pairs, $\left(G_\phi^P, D_\phi^P\right)$ and $\left(G_\phi^H, D_\phi^H\right)$ (Fig. 2). The G_ϕ^P (hologram-to-phase generator) transforms in-

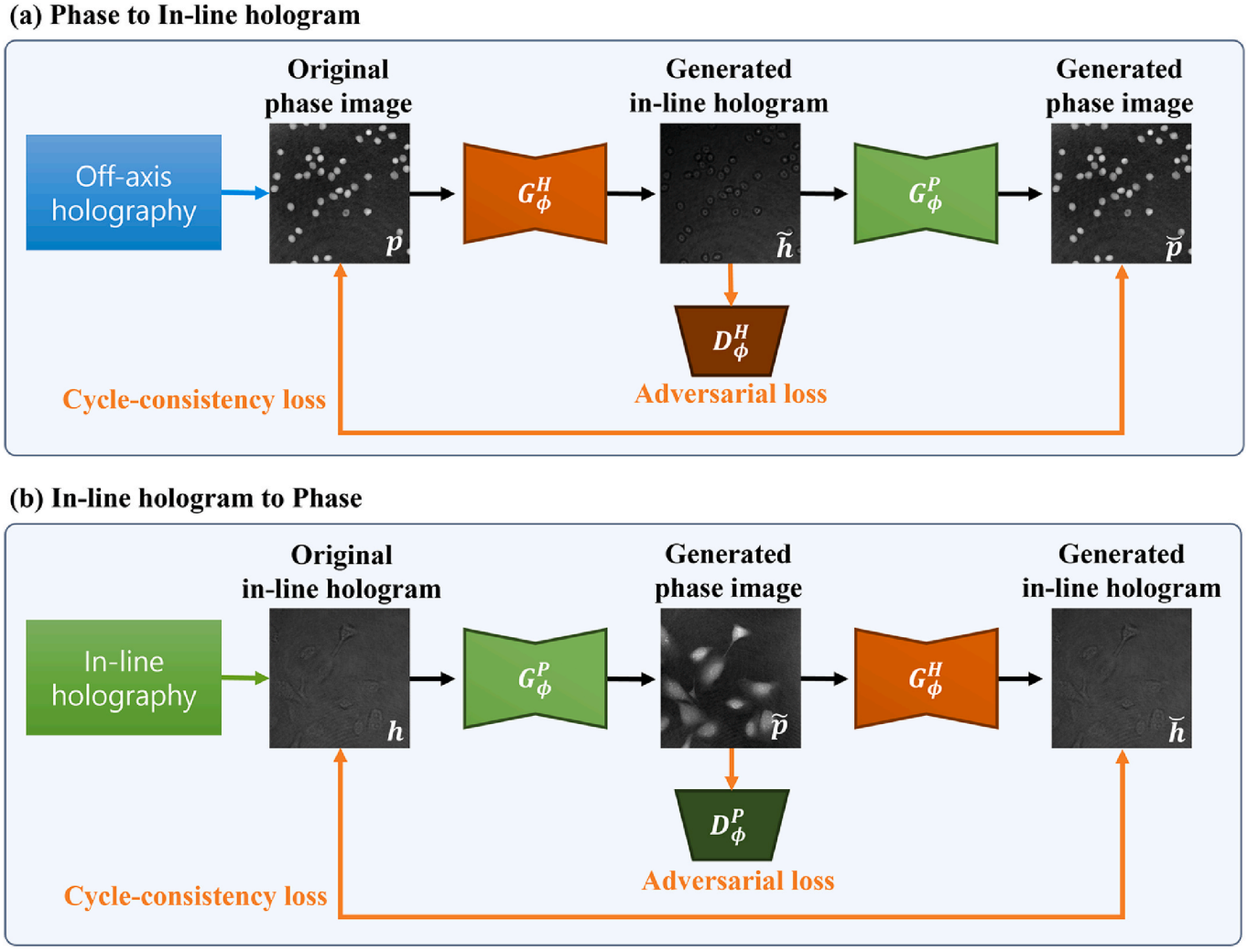


Fig. 3. Overview of the process for training the cycle-consistency module. The cycle-consistency module consists of two generators and two discriminators, enabling bidirectional image translation between two domains: the in-line hologram and the phase image. (a) The process of generating an in-line hologram using an original phase image and generating a corresponding phase image for cycle consistency learning. (b) The process of generating a phase image using an original in-line hologram and generating a corresponding in-line hologram for cycle consistency learning. The discriminators evaluate the quality of the generated phase images and in-line holograms by distinguishing them from real images.

line holograms into phase images, and the G_ϕ^H (phase-to-hologram generator) performs the reverse transformation (Zhu et al., 2017).

$$\tilde{p} = G_\phi^P(h) \tag{1}$$

$$\tilde{h} = G_\phi^H(p) \tag{2}$$

where, h and p represent the original in-line hologram and phase image, respectively, while $\tilde{p}$ and $\tilde{h}$ denote the phase image and in-line hologram generated by the cycle-consistency module. The generators of the cycle-consistency module used a ResNet backbone with two down-sampling, six residual, and two up-sampling blocks (He et al., 2016). The discriminators distinguish between generated images and real images for each domain. Discriminators help generators create images similar to actual images in each domain. The discriminators used five blocks with convolutional layer, batch normalization, and LeakyReLU.

To achieve unsupervised learning, the cycle-consistency module enforces round-trip consistency by transforming the generated phase images (by equation (1)) back to the in-line hologram domain and the generated in-line holograms (by equation (2)) back to the phase image domain through the corresponding generators (Zhu et al., 2017):

$$\breve{p} = G_\phi^P(\tilde{h}) \tag{3}$$

$$\breve{h} = G_\phi^H(\tilde{p}) \tag{4}$$

where, $\tilde{p}$ and $\tilde{h}$ denote the phase image and in-line hologram generated by the cycle-consistency module, while $\breve{p}$ and $\breve{h}$ are the corresponding reconstructions projected back to the original domain through the generator.

2.4.1.2. Denoising module. The denoising module represents the core reconstruction engine of the proposed framework, designed to generate high-quality images through an iterative refinement process in both directions: from in-line holograms to phase images and from phase images to in-line holograms. This bidirectional capability is essential for the overall framework to function effectively, as both transformations work synergistically to enable robust phase reconstruction from single-shot holographic measurements.

Fig. 4 shows the process of the denoising module. The module operates through two distinct processes: a forward process and a reverse process. The original phase images are obtained from off-axis holography, while the original in-line holograms are acquired from in-line holography. In the forward process, both phase images and in-line holograms are progressively corrupted by adding Gaussian noise over multiple timesteps following a predefined noise schedule, until they are transformed into pure random noise. This forward process systematically destroys the original image information by gradually increasing the noise level at each timestep. The reverse process learns to invert this corruption by training denoising networks to remove noise step by step, recovering the original information in each respective domain. Starting from pure noise, the reverse process iteratively refines the data through multiple denoising steps, gradually reconstructing either clean phase images or in-line holograms. This reverse process effectively learns the complex statistical relationships present in both holographic and phase data.

The forward process systematically corrupts clean images by

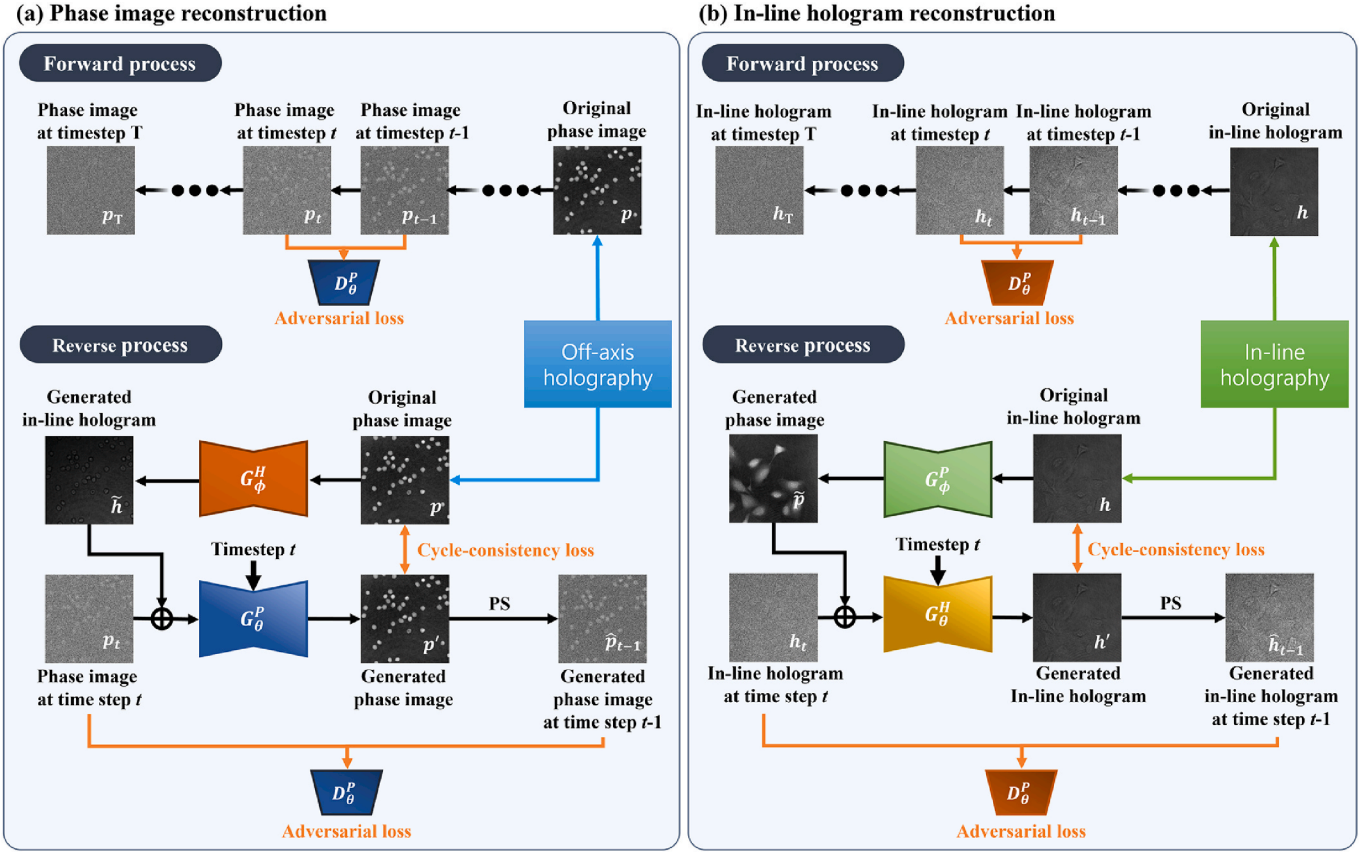


Fig. 4. Overview of the process for training the denoising module. The denoising module is designed to iteratively refine noisy images and reconstruct high-quality phase images through a step-by-step denoising process. This module comprises two sets of generator–discriminator pairs. (a) The process of reconstructing a phase image. (b) The process of reconstructing an in-line hologram. Training is divided into a forward process, where Gaussian noise is progressively added to the phase image, and a reverse process, where the model iteratively denoises the input to reconstruct the original phase information.

progressively adding Gaussian noise over T timesteps, as illustrated in Fig. 4. For phase image reconstruction (Fig. 4a), the process begins with a original phase image p and gradually transforms it through intermediate noisy states $p_{t\text{-}1}$, p_t until reaching pure Gaussian noise p_T at timestep T. Similarly, for in-line hologram reconstruction (Fig. 4b), the process starts with an original in-line hologram h and progresses through noisy intermediate states $h_{t\text{-}1}$, h_t to reach h_T.

The mathematical formulation of this corruption process follows the diffusion framework described by Ho et al.:

$$x_t^{(i,j)} = \sqrt{1-\beta_t}\,x_{t-1}^{(i,j)} + \sqrt{\beta_t}\,\varepsilon^{(i,j)},\ \varepsilon^{(i,j)} \sim N(0,1) \tag{5}$$

where, $t \in \{1,2,\ldots,T\}$, $x_t^{(i,j)}$ represents either the noisy phase image p_t or noisy in-line hologram h_t at timestep t and pixel coordinates (i, j), and $\varepsilon^{(i,j)}$ is Gaussian noise sampled independently for each pixel. In subsequent equations, the coordinates (i, j) will be omitted. The noise variance schedule β_t controls the amount of noise added at each timestep, following an exponentially scaled schedule (Song et al., 2020):

$$\beta_t = 1 - e^{-\overline{\beta}_{\min}\frac{1}{T} - 0.5(\overline{\beta}_{\max}-\overline{\beta}_{\min})\frac{2t-1}{T^2}} \tag{6}$$

where, $\overline{\beta}_{\min}$ and $\overline{\beta}_{\max}$ control the rate of noise variance increase through an exponentially scaled schedule. We set $T = 4$, $\overline{\beta}_{\min} = 0.1$, and $\overline{\beta}_{\max} = 20$.

The reverse process learns to systematically remove noise and reconstruct clean images from the corrupted data generated in the forward process. As illustrated in Fig. 4, this process operates in both directions: reconstructing phase images from noisy inputs (Fig. 4a) and reconstructing in-line holograms from noisy inputs (Fig. 4b).

The reverse process employs trainable neural networks to iteratively denoise images step by step. As shown in the figures, the denoising module comprises two sets of generator-discriminator pairs: (G_θ^P, D_θ^P) for phase image reconstruction and (G_θ^H, D_θ^H) for in-line hologram reconstruction. The generators take as input the noise-added image at timestep t along with conditioning information from the cycle-consistency module.

For phase image reconstruction (Fig. 4a), the generator G_θ^P receives the noisy phase image p_t and the in-line hologram $\widetilde{h}$ generated by the cycle-consistency module G_ϕ^H, producing a denoised phase image (Ho et al., 2020):

$$p' = G_\theta^P(p_t, \widetilde{h}, t) \tag{7}$$

Similarly, for in-line hologram reconstruction (Fig. 4b), the generator G_θ^H takes the noisy in-line hologram h_t and the phase image $\widetilde{p}$ from the cycle-consistency module G_ϕ^P (Ho et al., 2020):

$$h' = G_\theta^H(h_t, \widetilde{p}, t) \tag{8}$$

The generators employ U-Net architectures with seven downsampling blocks, one channel-wise self-attention block, and seven upsampling blocks (Ronneberger et al., 2015; Woo et al., 2018). A learnable temporal embedding corresponding to timestep t is incorporated as channel-specific bias into the feature maps within each sub-block to enable temporal conditioning.

To obtain the image at timestep t-1, posterior sampling (PS) is applied using the denoised predictions and the noisy input from the previous timestep (Kingma and Welling, 2013; Özbey et al., 2023; Xiao et al., 2021):

$$\widehat{x}_{t-1} = \overline{\mu}_t(\widehat{x}_t, x') + \overline{\beta}_t\varepsilon \tag{9}$$

$$\overline{\mu}_t(\widehat{x}_t, x') = \frac{\sqrt{\overline{\alpha}_{t-1}}\beta_t}{1-\overline{\alpha}_t}x' + \frac{\sqrt{\alpha_t}(1-\overline{\alpha}_{t-1})}{1-\overline{\alpha}_t}\widehat{x}_t, \overline{\beta}_t = \frac{1-\overline{\alpha}_{t-1}}{1-\overline{\alpha}_t}\beta_t \tag{10}$$

where $\widehat{x}_t$ is the generated in the previous timestep, respresented by $\widehat{h}_t$ or $\widehat{p}_t$ and x' is the denoised image, corresponding to h' or p'. $\widehat{x}_{t-1}$ is the image with one step of noise removed, resulting in $\widehat{h}_{t-1}$ or $\widehat{p}_{t-1}$ as the output of timestep t. $\alpha_t = 1-\beta_t$, $\overline{\alpha}_t = \prod_{r=[0,1,\ldots,t]}\alpha_r$ and $\varepsilon \sim N(0,1)$, which is sampled noise independently same as the forward process, except for the last denoising step where $\varepsilon = 0$.

The discriminator of the denoising module distinguishes between the t-1 image created from the model's predictions and calculations, and the t-1 image created by adding noise from the forward process. Each discriminator was composed of six sequential blocks, with each block containing two convolutional layers followed by a two-fold down-sampling operation. To enable temporal conditioning, a learnable temporal embedding was added to the feature maps within every block.

2.4.2. Model objective

The training objective operates through alternating discriminator and generator optimization stages, following the systematic process illustrated in the training flowchart (Fig. 5). The training begins with unpaired input data consisting of phase images p and in-line holograms h, which serve as the foundation for all subsequent computations.

Random timesteps t are independently selected for each domain, initiating the forward diffusion process that generates noisy sample pairs p_t, p_{t-1} and h_t, h_{t-1} according to Equation (5). Simultaneously, the generators within the cycle-consistency module establish cross-domain mappings: generator G_ϕ^H transforms phase image p into hologram $\widetilde{h}$ following Equation (2), while generator G_ϕ^P converts hologram h into phase image $\widetilde{p}$ using Equation (1). These cross-domain translations provide essential conditioning information for the subsequent denoising operations.

The generators within the denoising module then process the noisy inputs with cross-domain conditioning. Generator G_θ^P receives noisy phase image p_t along with cycle-generated hologram $\widetilde{h}$, producing denoised phase image p' through Equation (7). Concurrently, generator G_θ^H processes noisy hologram h_t with cycle-generated phase image $\widetilde{p}$, outputting denoised hologram h' via Equation (8). These denoised outputs undergo posterior sampling calculations following Equation (9) to generate $\widehat{p}_{t-1}$ and $\widehat{h}_{t-1}$, representing images with one timestep of noise removed.

During discriminator training (Fig. 5a), two distinct types of adversarial evaluation occur simultaneously. The discriminators within the cycle-consistency module assess the realism of cross-domain translations by distinguishing between real images and cycle-generated outputs. The hologram discriminator within the cycle-consistency module optimizes (Goodfellow et al., 2014):

$$L_{D_\phi^H} = \mathbb{E}\left[-log\left(D_\phi^H(h)\right)\right] + \mathbb{E}\left[-log\left(1-D_\phi^H(\widetilde{h})\right)\right] \tag{11}$$

while the phase discriminator within the cycle-consistency module optimizes:

$$L_{D_\phi^P} = \mathbb{E}\left[-log\left(D_\phi^P(p)\right)\right] + \mathbb{E}\left[-log\left(1-D_\phi^P(\widetilde{p})\right)\right] \tag{12}$$

The discriminators within the denoising module evaluate the denoising progression by comparing authentic temporal sequences against generated ones. The hologram discriminator within the denoising module distinguishes between real progression h_t, h_{t-1} and generated progression h_t, $\widehat{h}_{t-1}$ (Özbey et al., 2023):

$$L_{D_\theta^H} = \mathbb{E}\left[-log\left(D_\theta^H(t,h_t,h_{t-1})\right)\right] + \mathbb{E}\left[-log\left(1-D_\theta^H(t,h_t,\widehat{h}_{t-1})\right)\right] \tag{13}$$

Similarly, the phase discriminator within the denoising module optimizes:

$$L_{D_\theta^P} = \mathbb{E}\left[-log\left(D_\theta^P(t,p_t,p_{t-1})\right)\right] + \mathbb{E}\left[-log\left(1-D_\theta^P(t,p_t,\widehat{p}_{t-1})\right)\right] \tag{14}$$

Generator training follows identical forward processing but focuses

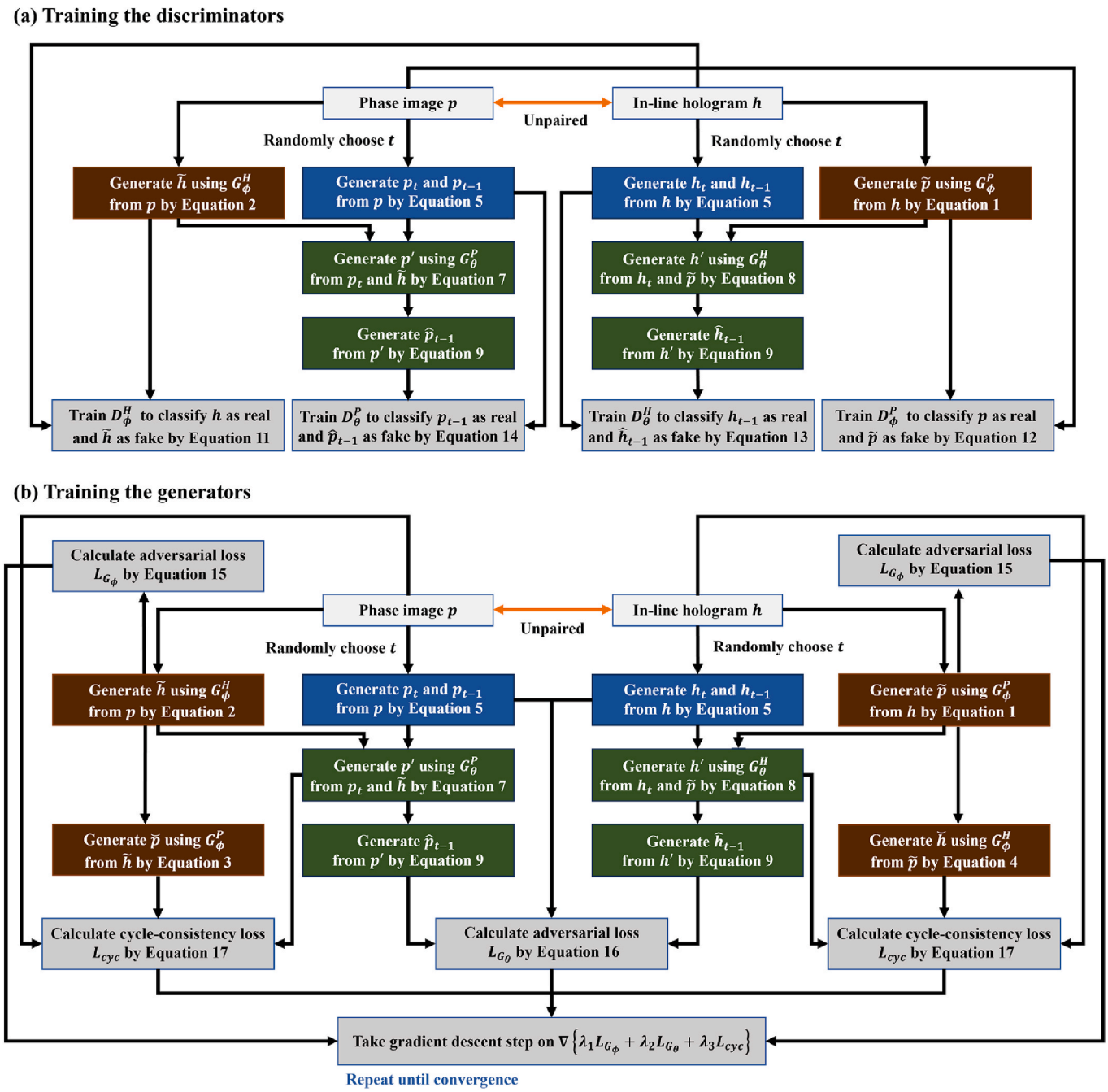


Fig. 5. Flowchart illustrating the training process of the unsupervised diffusion model for holographic phase reconstruction. (a) Training the discriminators and (b) training the generators through alternating optimization stages. The light gray boxes represent the unpaired input data: phase images p and in-line holograms h, which do not require direct correspondence. The blue boxes indicate the forward process that progressively adds noise to generate p_t and h_t at timestep t. The green boxes represent the reverse process that iteratively removes noise to reconstruct clean images. The brown boxes denote the cycle-consistency module operations that establish bidirectional mappings between hologram and phase domains. The gray boxes show objective function calculations including adversarial losses and cycle-consistency loss. The training alternates between discriminator optimization (a) and generator optimization (b) through gradient descent steps, repeating until convergence. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

on optimizing generator performance against the trained discriminators (Fig. 5b). The generators within the cycle-consistency module receive adversarial feedback designed to fool their respective discriminators within the cycle-consistency module (Goodfellow et al., 2014):

$$L_{G_\phi} = \mathbb{E}\left[-log\left(D_\phi^H(\tilde{h})\right)\right] + \mathbb{E}\left[-log\left(D_\phi^P(\tilde{p})\right)\right] \tag{15}$$

The generators within the denoising module optimize their adversarial performance against the discriminators within the denoising module, aiming to produce temporally consistent denoising progressions (Özbey et al., 2023):

$$L_{G_\theta} = \mathbb{E}\left[-log\left(D_\theta^P(t, p_t, \hat{p}_{t-1})\right)\right] + \mathbb{E}\left[-log\left(D_\theta^H(t, h_t, \hat{h}_{t-k})\right)\right] \tag{16}$$

The cycle-consistency constraint provides crucial physical consistency by measuring reconstruction errors across multiple pathways. The cycle reconstruction pathway evaluates round-trip translations where phase images p are converted to holograms $\tilde{h}$ and back to phase images $\breve{p}$, while holograms h undergo conversion to phase images $\tilde{p}$ and back to

holograms h. The consistency with denoised outputs ensures that the generators within the denoising module maintain fidelity to the original inputs, where h' represents the denoised hologram output from the denoising module's generator G_θ^H and p' represents the denoised phase image output from the denoising module's generator G_θ^P.The complete cycle-consistency loss encompasses (Özbey et al., 2023; Zhu et al., 2017):

$$L_{cyc} = \mathbb{E}\left[\left(|p - \breve{p}|_1 + \left|h - \breve{h}\right|_1\right) + \left(|p - p'|_1 + |h - h'|_1\right)\right] \tag{17}$$

The integrated training objective combines these components through weighted loss terms. Generator training optimizes:

$$L_G^{total} = \lambda_1 L_{G_\phi} + \lambda_2 L_{G_\theta} + \lambda_3 L_{cyc} \tag{18}$$

while discriminator training minimizes:

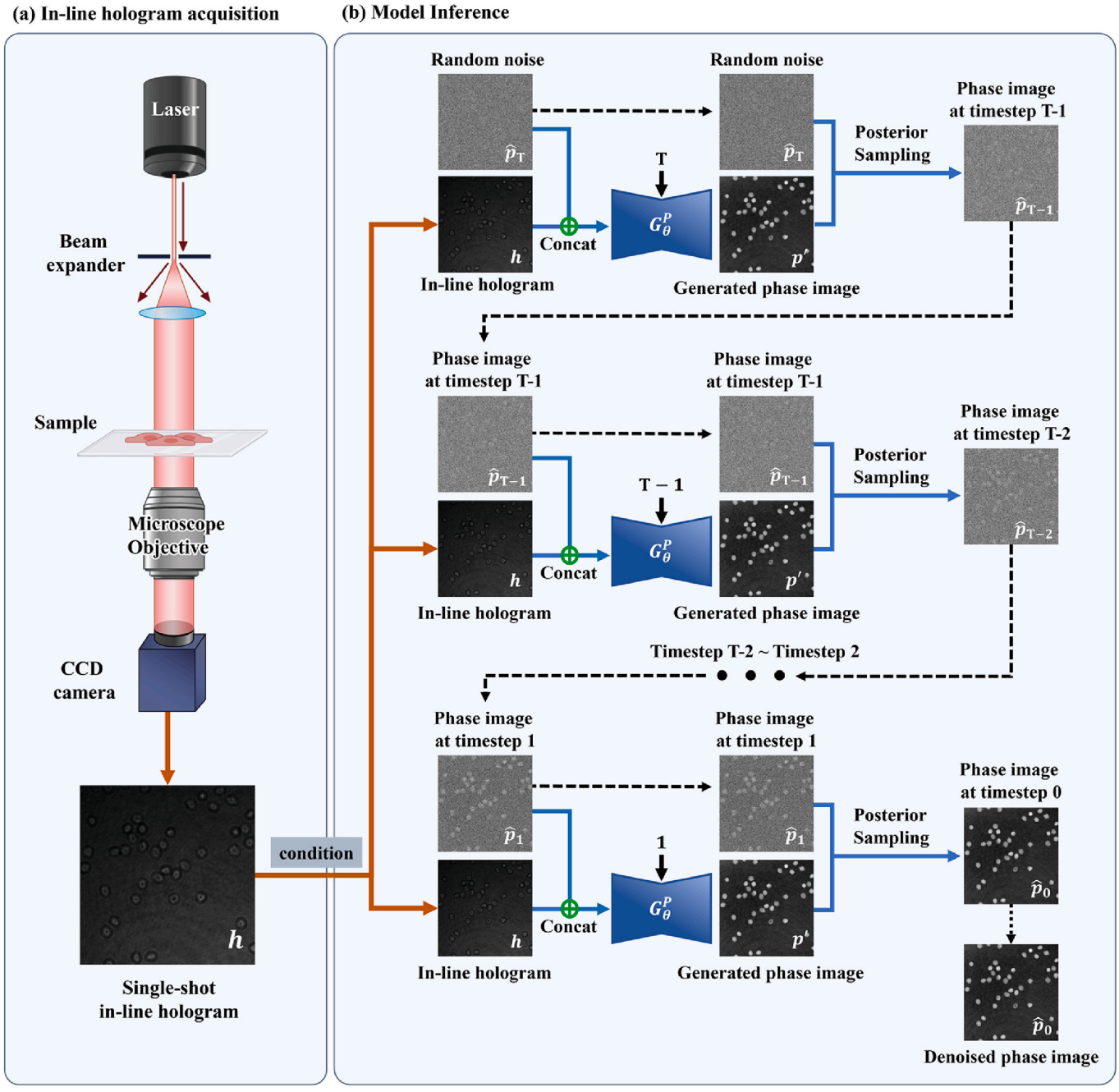


Fig. 6. Inference process that generates a phase image from a single in-line hologram and Gaussian random noise sample. (a) In-line hologram acquisition using a simplified optical setup where the sample is directly illuminated by laser light through a microscope objective and recorded by a CCD camera, producing a single-shot in-line hologram that serves as conditioning input for the model. (b) Model inference showing the iterative denoising process from timestep T to timestep 0. The trained generator G_θ^P progressively denoises random Gaussian noise $\hat{p}_T$ by conditioning on the input in-line hologram h through concatenation operations. At each timestep, the generator produces a denoised phase image p', which undergoes posterior sampling to generate $\hat{p}_{t-1}$ with one step of noise removed. This iterative process continues until timestep 0, producing the final denoised phase image $\hat{p}_0$ that effectively resolves the twin image problem inherent in the original in-line hologram.

$$L_D^{total} = \lambda_1 \left(L_{D_\phi^P} + L_{D_\phi^H} \right) + \lambda_2 \left(L_{D_\theta^P} + L_{D_\theta^H} \right) \quad (19)$$

The training alternates between these stages through gradient descent optimization with loss weights $\lambda_1 = \lambda_2 = 1$ and $\lambda_3 = 0.5$, ensuring balanced adversarial training while maintaining cycle-consistency constraints essential for robust holographic phase reconstruction.

2.4.3. Model inference

Once training is completed, the model operates as a true single-shot system that requires only an in-line hologram to reconstruct quantitative phase images, as illustrated in Fig. 6. This represents a fundamental departure from the training requirement where off-axis holograms were necessary to provide reference phase information. The critical distinction is that off-axis holography is used exclusively during the one-time training process to establish reliable phase reconstruction capabilities, after which the trained model can process any new in-line hologram without requiring additional off-axis measurements.

The inference process begins with a single captured in-line hologram h acquired through the simplified optical setup shown in Fig. 6a and random Gaussian noise $\hat{p}_T$ initialized at timestep T, as demonstrated in the inference flowchart (Fig. 6b). The trained generator G_θ^P within the denoising module processes the concatenated input consisting of the noisy phase sample $\hat{p}_t$ and the conditioning in-line hologram h, producing a denoised phase image p' through Equation (7). This denoised output undergoes posterior sampling following Equation (9) to generate $\hat{p}_{t-1}$, representing a phase image with one timestep of noise removed. The timestep counter is decremented ($t \leftarrow t$-1), and the iterative denoising process continues through multiple timesteps as illustrated in Fig. 6b until reaching timestep 0, where the final denoised phase image $\hat{p}_0$ is obtained. The flow chart for the inference process is shown in Fig. 7.

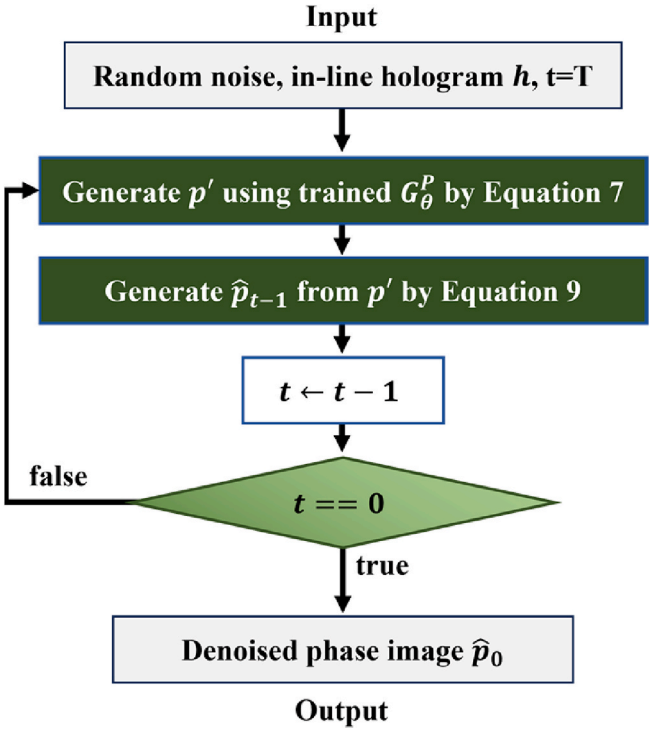


Fig. 7. Flowchart depicting the inference process of the trained unsupervised diffusion-based model for quantitative phase image reconstruction. Starting from a single acquired in-line hologram and an initial Gaussian noise sample, the model iteratively performs denoising through reverse diffusion steps, progressively reconstructing a high-quality quantitative phase image. The inference continues iteratively from timestep T down to timestep 0, achieving a precise final phase image reconstruction.

This iterative denoising process leverages the trained model's learned understanding of the relationship between in-line holographic patterns and their corresponding phase distributions. The model applies its learned knowledge of holographic physics and twin image suppression without requiring any additional measurements or off-axis reference data. Each denoising step progressively removes noise while conditioning on the input hologram, ensuring that the final reconstruction maintains consistency with the measured holographic data.

The final output $\hat{p}_0$ provides a quantitative phase image that has effectively resolved the twin image problem inherent in the original in-line hologram. This reconstructed phase image contains the optical path difference information necessary for biological analysis, including cellular morphology, dry mass calculation, and 3D structure characterization, all derived from a single holographic measurement without requiring the complex optical infrastructure traditionally associated with quantitative phase imaging.

2.5. Evaluation metrics

To more thoroughly assess the visual fidelity of the image translation results, we utilized two standard metrics commonly applied in image-to-image translation tasks. The first metric, peak signal-to-noise ratio (PSNR), measures the pixel-wise intensity difference between the reconstructed image and the ground truth, with higher values indicating greater similarity. The second metric, structural similarity index (SSIM) (Wang et al., 2004), evaluates perceptual similarity by considering luminance, contrast, and structural components. An SSIM value approaching 1 signifies a higher degree of resemblance between the two images.

3. Results

3.1. Phase reconstruction using the diffusion model

3.1.1. Results of the diffusion model that trained only red blood cells or cancer cells

We trained a diffusion model using in-line holograms and phase images of only RBCs or cancer cells. In the RBC data we used, both discocytes and spherocytes existed simultaneously (Kim et al., 2022). The phase images generated by the diffusion model were evaluated against ground truth phase images reconstructed from off-axis holograms. Fig. 8 presents both the reconstructed phase maps and corresponding 3D surface plots, demonstrating that our method successfully preserves the 3D structural information inherent in holographic imaging. The 3D surface plots clearly reveal the characteristic biconcave shape of red blood cells and the complex morphological features of cancer cells, confirming that the depth information is accurately reconstructed by our approach. The left column depicts results from a model trained solely on RBCs, while the right column shows outcomes from training exclusively on cancer cells.

As illustrated in Fig. 8, the phase line profiles extracted from the cropped regions demonstrate strong agreement between the diffusion model output and the ground truth, providing quantitative evidence that our method maintains the same 3D structural characteristics as the reference measurements. These line profiles show that the phase variations across cellular features, which directly correspond to optical path differences and thus depth information, are accurately reproduced by our model, confirming that the 3D morphological information is faithfully reconstructed.

The reconstruction performance shows some differences between the two cell types due to their distinct morphological characteristics. RBCs, with their consistent round morphology and uniform structural features, show excellent agreement with ground truth reconstructions. Cancer cells, however, tend to exhibit slightly blurred edges in the reconstructed phase images, particularly at cellular boundaries and extended processes. This occurs because cancer cells often display complex

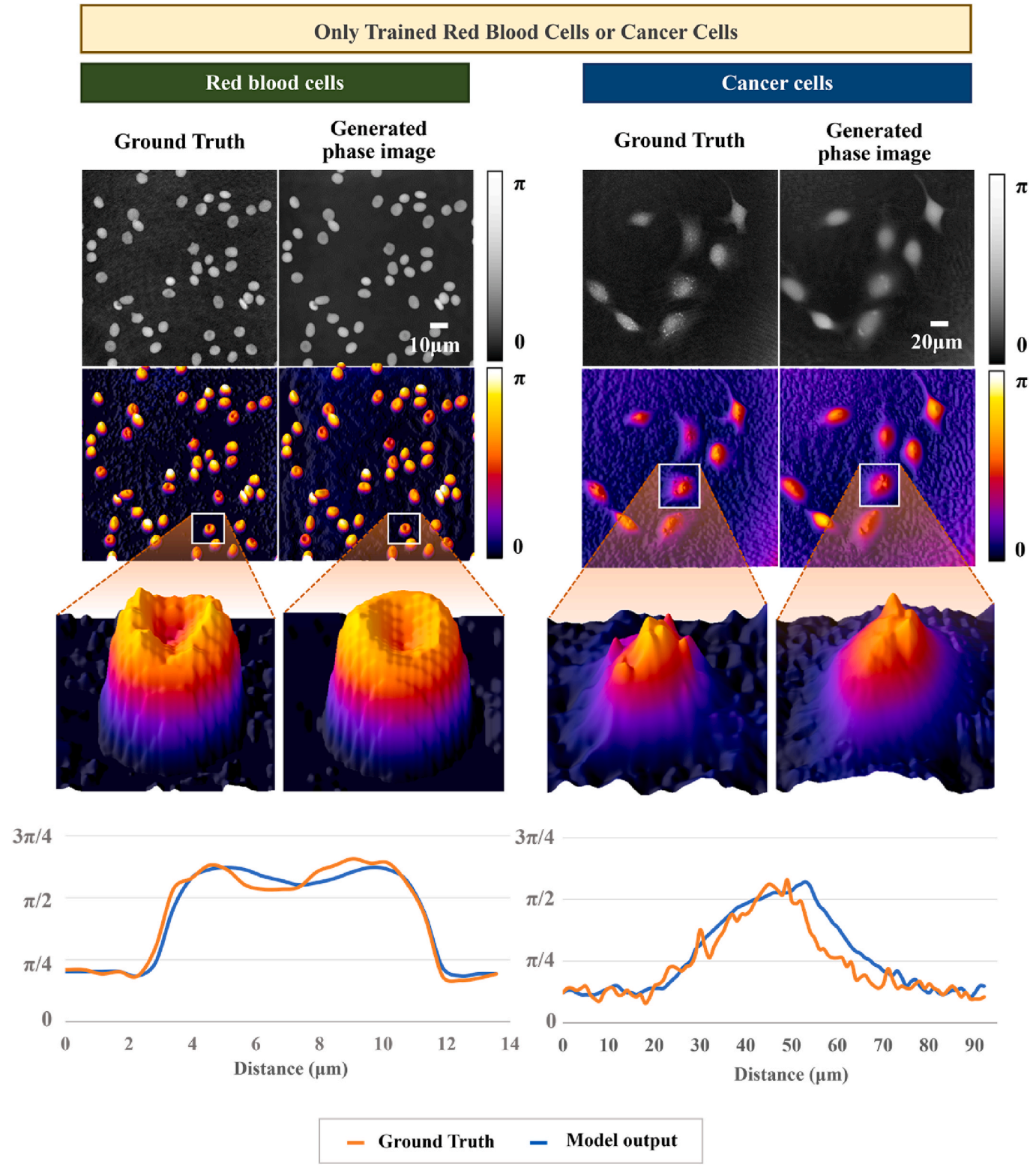


Fig. 8. Red blood cell and cancer cell phase images generated from a diffusion model only trained on red blood cell or cancer cell holographic images. The ground truth image is shown on the left, and the corresponding image generated by the model is shown on the right. Below, a 3D surface plot of the reconstructed phase image is presented using a lookup table, followed by a magnified 3D view of the region marked with a white box. The bottom graphs display line profiles of the phase distribution across representative cells, comparing ground truth (orange) and generated results (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

branching structures and irregular extensions that present greater challenges for reconstruction compared to the more uniform geometry of RBCs. The consistent morphology among RBCs makes them more amenable to accurate reconstruction, whereas the diverse and complex morphological features of cancer cells, including their extended projections and irregular boundaries, can result in some edge blurring in the reconstructed phase images.

3.1.2. Results of the diffusion model that simultaneously trained red blood cells and cancer cells

Effective training of a deep learning model for phase reconstruction requires strong generalization capability to ensure robust performance across diverse samples. Even if data of different shapes are used as input, the performance of phase reconstruction should be the same. We trained a diffusion model using simultaneous in-line holographic and phase images of RBCs and cancer cells. The output phase image produced by the diffusion model was compared with the ground truth phase image obtained through numerical reconstruction of the off-axis hologram. Fig. 9 shows phase images and corresponding 3D surface plots generated from the diffusion model, demonstrating that our method continues to preserve the essential 3D structural information when trained on multiple cell types simultaneously. The 3D surface plots clearly show that both the characteristic biconcave morphology of red blood cells and the complex structural features of cancer cells are accurately reconstructed, confirming that the depth and dimensional information inherent in holographic measurements is maintained across different cellular types.

Fig. 9 shows comparable results to those presented in Fig. 8, indicating that simultaneous training on multiple cell types does not compromise the reconstruction quality. The phase line profiles generated by the diffusion model exhibit a high degree of similarity to the reference measurements, providing quantitative evidence that the 3D morphological characteristics are preserved regardless of the training strategy. These profiles demonstrate that the phase variations corresponding to optical path differences and cellular depth information are accurately reproduced across both cell types, validating that our method successfully maintains 3D reconstruction capabilities when handling diverse biological specimens simultaneously.

The reconstruction performance maintains the same characteristics observed in single-cell-type training: RBCs continue to show excellent reconstruction fidelity due to their consistent morphology, while cancer cells may exhibit slight edge blurring at cellular boundaries and extended processes due to their more complex and variable structural features. However, the overall 3D reconstruction quality remains high for both cell types, demonstrating the robustness of our approach across diverse cellular morphologies.

3.1.3. Single cell analysis of phase images reconstructed from the diffusion model

The reconstructed phase image offers detailed quantitative insights into cellular morphology. Based on this image, we computed the optical path difference (OPD) (Rappaz et al., 2005; Roitshtain et al., 2017). The OPD is directly proportional to the physical thickness of the cell and serves as a key parameter for quantitative phase analysis, providing essential 3D information about cellular structure and volume. To analyze individual cells, we applied image smoothing followed by Otsu thresholding to generate binary masks, which were then used to segment single-cell regions (Fig. 10a). For each segmented cell, we extracted three key 3D metrics: dry mass, projected area, and phase volume. Dry mass reflects the total non-aqueous content within the cell; the projected area represents the cell's two-dimensional footprint; and phase volume offers an OPD-based estimation of the intracellular volume, accounting for variations in refractive index and cellular thickness.

These quantitative measurements demonstrate that our reconstructed phase images preserve the essential 3D information necessary for accurate cellular analysis. The successful calculation of phase volume, in particular, validates that our method maintains the depth-related information encoded in the original holographic measurements, as this parameter directly depends on the optical path differences that reflect the 3D cellular structure.

Fig. 10 compares the quantitative measurements of the ground-truth phase images with those obtained from the phase images reconstructed by each model. When comparing the extracted information, the diffusion model shows almost similar results whether it trains only RBCs or cancer cells, or simultaneously trains both data. The strong correlation between measurements from ground truth and reconstructed phase images for all 3D parameters demonstrates that our method successfully preserves the quantitative 3D information necessary for accurate cellular morphometry and volumetric analysis. The reconstructed results show strong agreement with the ground truth for RBCs, owing to their consistent shape, whereas cancer cells exhibit minor deviations due to their structural heterogeneity.

3.2. Comparison results with GAN models based on unsupervised learning

3.2.1. Comparison results that trained only red blood cells or cancer cells

We compared the results of phase reconstruction of the diffusion model with CycleGAN and UNIT, which are famous unsupervised learning-based GAN models (Fig. 11). CycleGAN shows poor reconstruction, with dark regions of the inline hologram barely being reconstructed. For the UNIT model, the phase image appears to be well reconstructed for RBCs, but there are parts that appear blurry. Unsupervised GAN models struggle with cancer cell reconstruction, displaying significant shape variances from the ground truth and consistently missing numerous cells. This challenge is fueled by the vastly diverse morphology of cancer cells, rendering precise image reconstruction a quite difficult task.

3.2.2. Comparison results that simultaneously trained red blood cells and cancer cells

Unlike the previous results, unsupervised GAN-based models, when trained simultaneously on two different types of data, show very large problems in phase reconstruction (Fig. 11). In the case of CycleGAN, it can be seen in red blood cell reconstruction that an image similar to a cancer cell is reconstructed. Additionally, in cancer cell reconstruction, no image is reconstructed at all. In the case of UNIT, while in RBC reconstruction, some cells are reconstructed to some extent, many other cells are missing. In cancer cell reconstruction, the shape is barely recognizable. In contrast, the diffusion model generated images similar to the ground truth, regardless of RBCs and cancer cells. This shows that the model can simultaneously train more diverse cells, and generalization is possible.

3.2.3. Evaluation metrics for diffusion and comparative models

Tables 1 and 2 present the outcomes of evaluating the similarity between the generated images and the ground truth, using SSIM and PSNR as key metrics. In the context of red blood cells (RBCs), GAN-based models demonstrated a PSNR of around 30, whereas the diffusion model exhibited a notably higher PSNR exceeding 33. Similarly, in SSIM calculations, the model recorded the highest values. Interestingly, when trained exclusively on RBCs, the performance of the GAN-based models improved, compared to when simultaneous trained on RBCs and cancer cells. However, in the diffusion model, this trend was reversed, which model showed significantly better results when concurrently trained on both RBCs and cancer cells. A parallel pattern emerged with cancer cells: when trained solely on cancer cells, the GAN-based models achieved superior outcomes, as opposed to when trained jointly with RBCs. This differential was more pronounced than in the RBC test. In contrast, the diffusion model maintained high-performance indices, regardless of whether it was trained only on cancer cells, or simultaneously on both cell types. These findings suggest that for the GAN-based models, simultaneous training of different data types may lead to confusion in distribution learning. Conversely, for the diffusion model, training on a

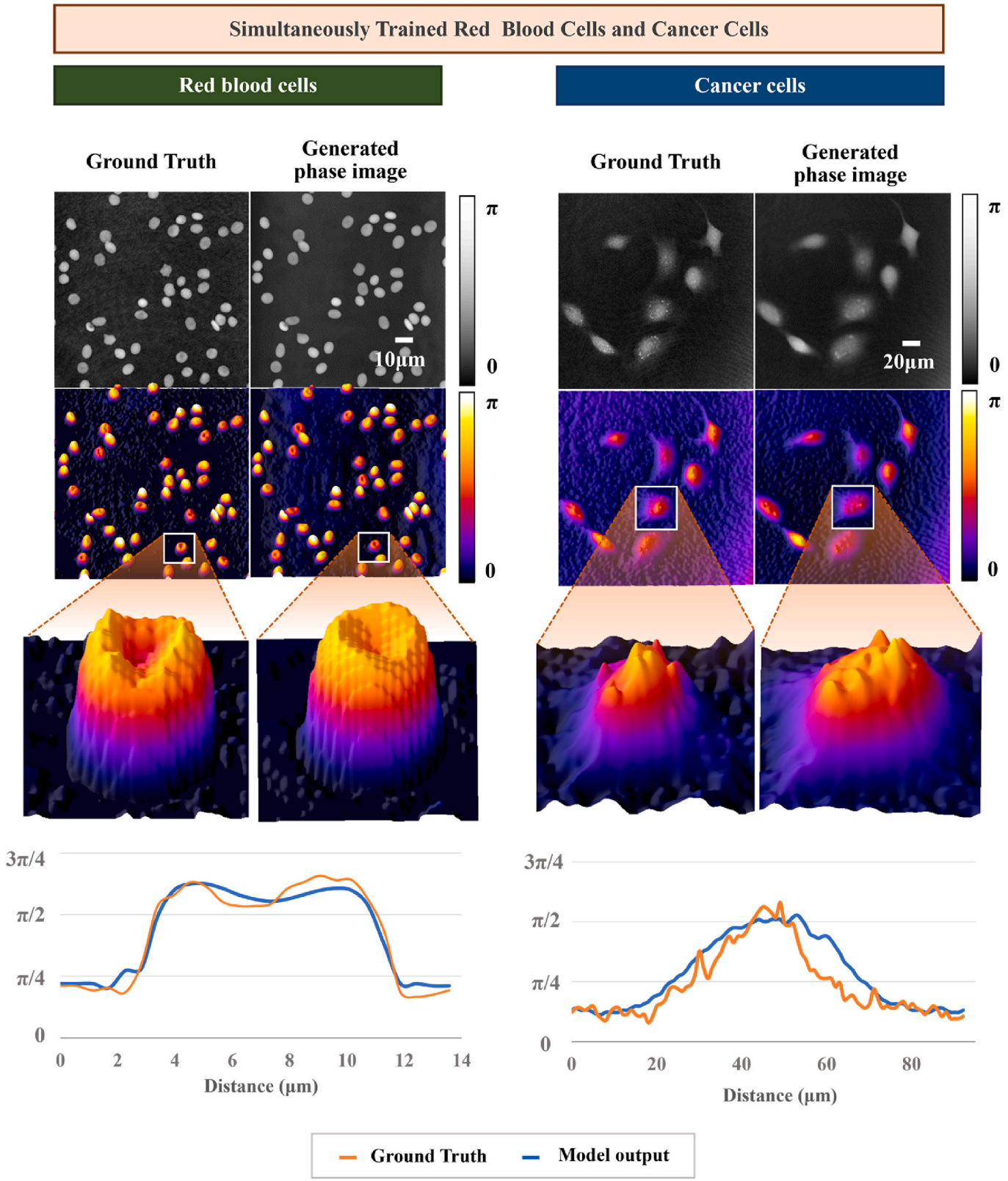


Fig. 9. Red blood cell and cancer cell phase images generated from a diffusion model simultaneously trained on red blood cell and cancer cell holographic images. The ground truth image is shown on the left, and the corresponding image generated by the model is shown on the right. Below, a 3D surface plot of the reconstructed phase image is presented using a lookup table, followed by a magnified 3D view of the region marked with a white box. The bottom graphs display line profiles of the phase distribution across representative cells, comparing ground truth (orange) and generated results (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

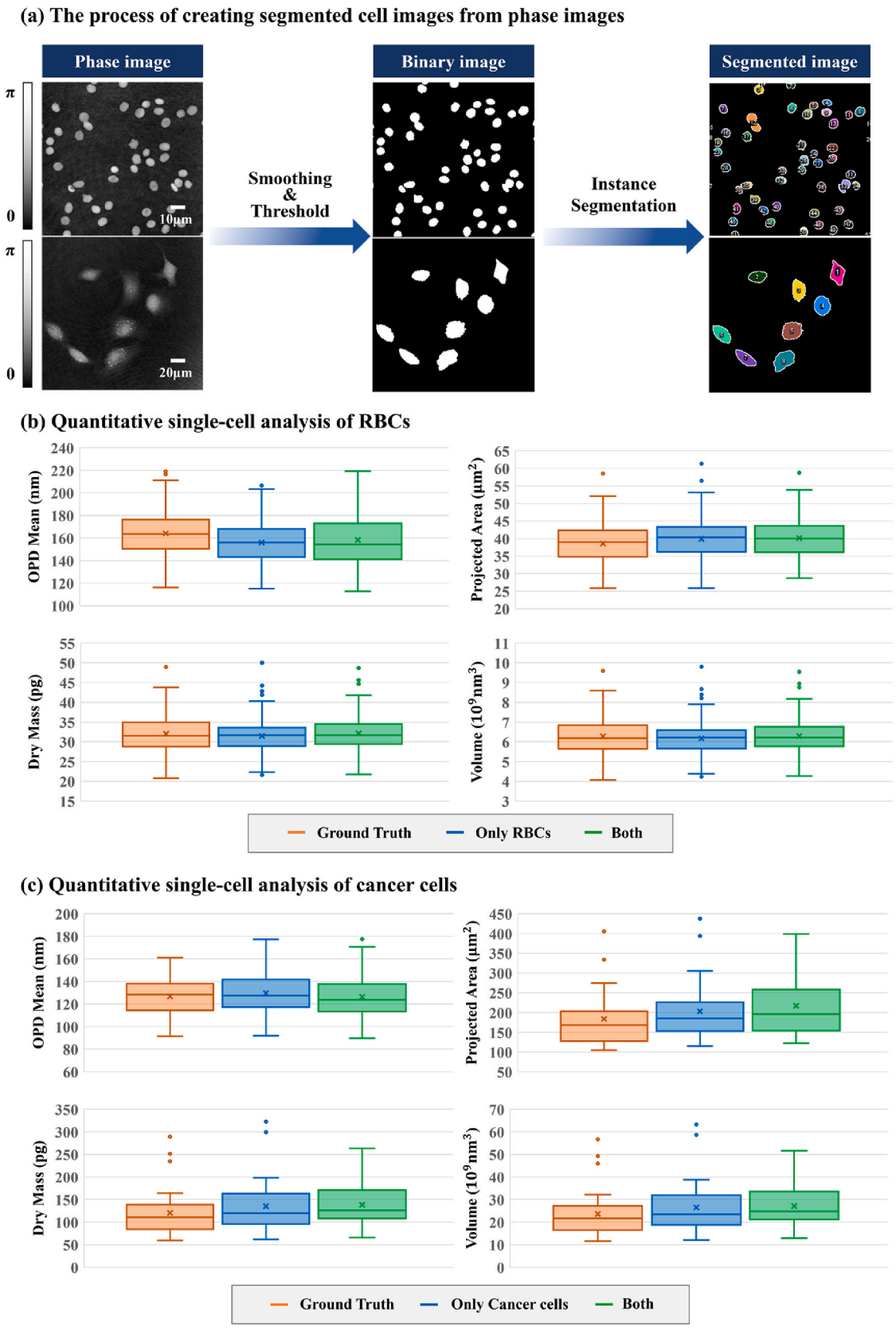


Fig. 10. Quantitative analysis of cellular features extracted from phase images generated by the diffusion model. (a) is the process of creating segmented cell images for cell analysis. (b) is single cell analysis for RBCs data, and (c) is single cell analysis for cancer cell data. Four indicators were calculated from cell images. The orange box is the ground truth, while the blue box results from a model trained only RBCs or cancer cells. The green box results from a model simultaneously trained on RBCs and cancer cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

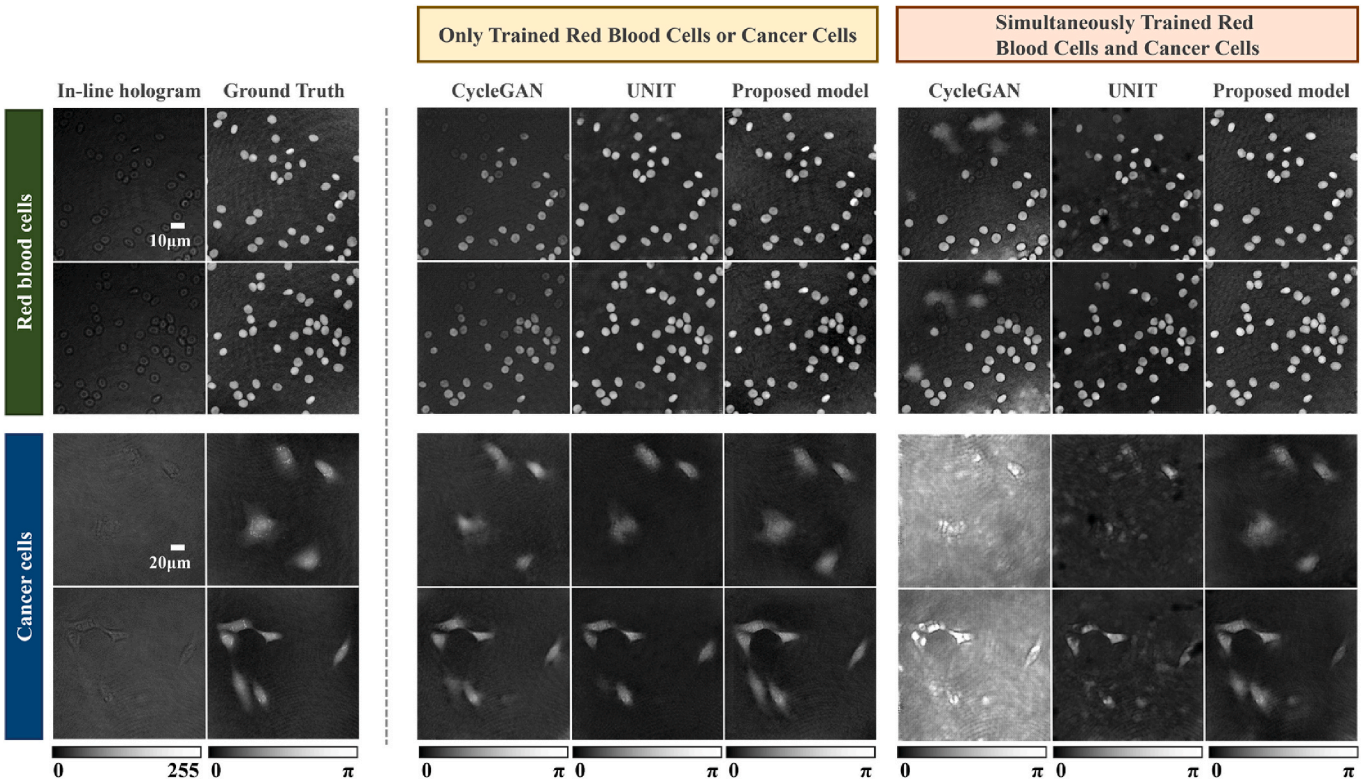


Fig. 11. Results of the red blood cell and cancer cell phase image generated by CycleGAN, UNIT, and the diffusion model. The left side of the result is comparison results that trained only red blood cells or cancer cells, and the right side of the result is comparison results that simultaneously trained red blood cells or cancer cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
PSNR and SSIM values for phase images generated by models trained only red blood cells or cancer cells from 200 test images (left: PSNR, right: SSIM).

Model	Red blood cells	Cancer cells
CycleGAN	31.55/0.7253	32.07/0.8255
UNIT	31.59/0.7402	31.34/0.8029
Diffusion model	33.09/0.7625	33.90/0.8449

Table 2
PSNR and SSIM values for phase images generated by models simultaneously trained on red blood cells and cancer cells from 200 test images (left: PSNR, right: SSIM).

Model	Red blood cells	Cancer cells
CycleGAN	29.12/0.5384	28.73/0.4999
UNIT	30.96/0.7318	31.27/0.7506
Diffusion model	34.26/0.8359	35.27/0.8591

diverse data set appears to facilitate more robust feature extraction.

3.2.4. *Computational cost for diffusion and comparative models*

While our diffusion model demonstrates superior reconstruction performance compared to CycleGAN and UNIT, it is essential to evaluate the computational cost associated with these improvements. We performed our comparison using a single NVIDIA Quadro RTX 6000. As shown in Table 3, the diffusion model has significantly more parameters compared to CycleGAN and UNIT, representing approximately 17-fold and 16-fold increases respectively. This larger model size translates to higher memory requirements compared to CycleGAN and UNIT, requiring approximately 17 times more memory resources. Additionally, the training time per epoch is substantially longer for the diffusion model (3.5275 s) compared to CycleGAN (0.40375 s) and UNIT (0.49625 s), requiring approximately 9 times more computational resources during training. These increased computational demands during training reflect the complex iterative denoising process and the dual-module architecture incorporating both cycle-consistency and denoising components, which contribute to the model's enhanced reconstruction capabilities.

Table 3
Model complexity and computational efficiency comparison: parameters, memory usage, training time, and inference time for diffusion model, CycleGAN, and UNIT.

Model	# Parameters	Memory usage	Training time	Inference time
CycleGAN	21,194,116	80.86 MB	0.403 s	0.025 s
UNIT	22,327,684	85.18 MB	0.496 s	0.058 s
Diffusion model	361,260,426	1378.1 MB	3.527 s	0.271 s

Despite the higher training costs, the diffusion model demonstrates practical efficiency during inference with a reconstruction time of 0.271 s per image, which approaches real-time performance suitable for clinical applications. While this is not faster than CycleGAN (0.025 s) or UNIT (0.058 s), the inference speed remains acceptable for practical deployment, especially considering the substantial quality improvements achieved. The reconstruction quality gains, as demonstrated by consistently higher PSNR and SSIM values across both red blood cells and cancer cells, justify the computational overhead. The trade-off analysis reveals that while the diffusion model requires more computational resources during training and slightly longer inference times, the superior reconstruction fidelity and robustness across diverse cellular morphologies provide significant value for quantitative phase imaging applications. This computational cost-to-performance ratio represents a worthwhile investment for applications where

reconstruction accuracy is paramount, such as medical diagnostics and biological research.

3.3. Comparison results with supervised learning-based model in scenarios with limited data

3.3.1. Comparison results that trained only red blood cells or cancer cells

We compared the phase reconstruction results of the unsupervised diffusion model with that of the supervised model (Fig. 12). The supervised model was trained on paired data using only the denoising module, which excludes the cycle consistency module incorporated in the unsupervised model. The comparison was also conducted in a scenario characterized by a scarcity of labeled data, with each model being trained using merely 50 labels.

When trained with only RBCs, both the supervised model and the diffusion model showed excellent performance when all available data was used for training. When limited to training using only 50 labels, the supervised model showed that certain cells were not reconstructed, but overall the reconstruction was performed well. The diffusion model reliably reconstructed cells. When only cancer cells were trained, a similar trend was observed as for RBCs, but a larger difference occurred. In particular, when limited to 50 labels, the supervised model showed a very large difference from the ground truth. This can be inferred as a result of the diversity of cell morphology. Because most cells in RBCs have similar morphology, a supervised model can be trained to perform reconstruction even if only a small amount of data is used. However, because cancer cells vary significantly in their morphology, using a small amount of labeled data can lead to a decrease in performance. The unsupervised learning model shows better generalization compared to the supervised learning model, as it can efficiently utilize a large amount of input data and train across the domain.

3.3.2. Comparison results that simultaneously trained red blood cells and cancer cells

Similar to previous results, we compared the phase reconstruction results of the diffusion model with those of the supervised model when simultaneously training on both red blood cells and cancer cells (Fig. 12). When utilizing all available data, the supervised model effectively reconstructed the phase images of both red blood cells and cancer cells. However, a significant decline in performance was observed when each model was trained using only 50 labels for each cell type. In the case of cancer cells, the difference was more pronounced, and a noticeable deterioration in the reconstruction quality of red blood cells was also evident. In contrast, the diffusion model demonstrated the capability to generate images similar to the ground truth, both when using the full dataset and when limited to 50 labels per cell type.

3.3.3. Evaluation metrics for diffusion and comparative models

Tables 4 and 5 present the results of evaluating the similarity between generated images and the ground truth using SSIM and PSNR. The supervised model exhibits excellent performance when trained with the full dataset, particularly when simultaneously training from both red blood cells and cancer cells, showing high SSIM and PSNR values for both cell types. However, the model's performance declines when the number of labels is limited to 50. This performance drop is particularly pronounced in the phase reconstruction of cancer cells, where both SSIM and PSNR values significantly decrease. In contrast, the diffusion model maintains consistent performance regardless of the number of labels. In particular, it shows almost no decrease in performance in the reconstruction of cancer cell phase images, where the performance of the

Table 4
PSNR and SSIM values for phase images generated by models trained only red blood cells or cancer cells from 200 test images (left: PSNR, right: SSIM).

Model	Red blood cells	Cancer cells
Supervised model	32.81/0.8862	35.64/0.9091
Supervised model with 50 labels	32.46/0.8668	29.74/0.8046
Diffusion model	33.09/0.7625	34.90/0.8449
Diffusion model with 50 labels	31.99/0.7291	33.01/0.8345

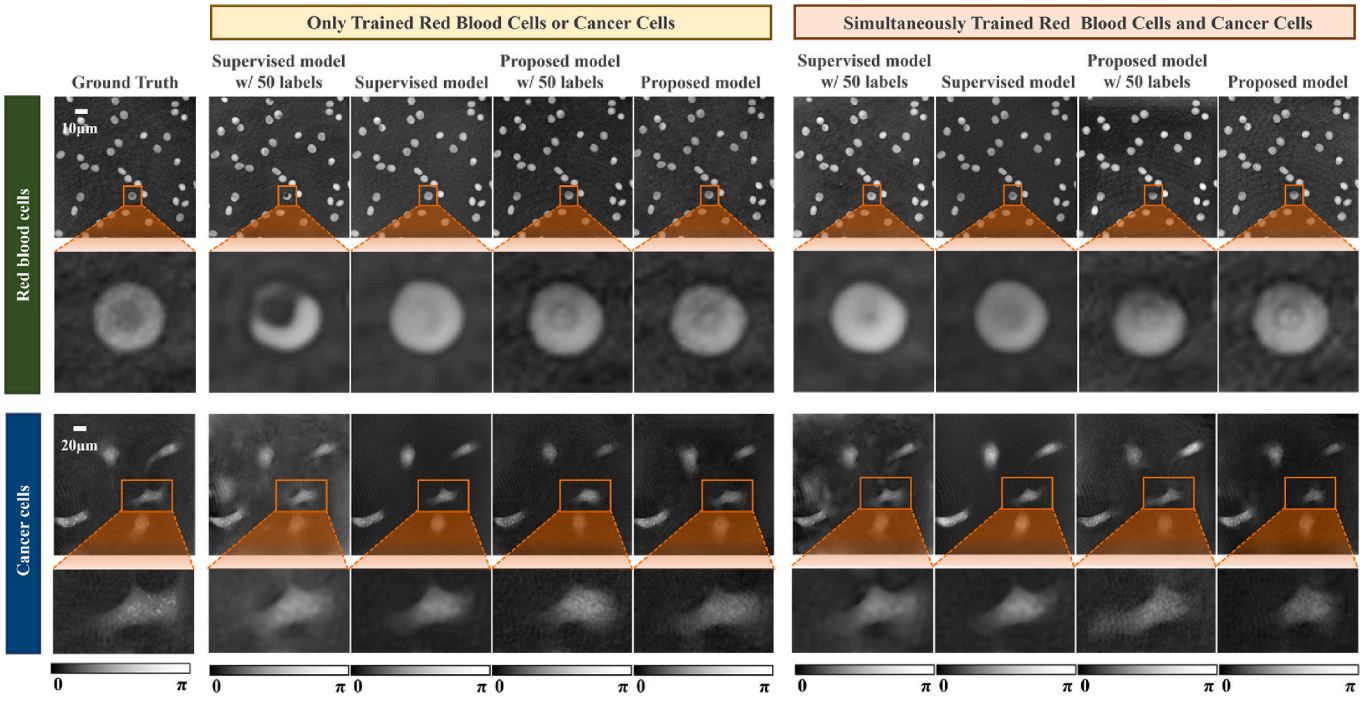


Fig. 12. Results of the red blood cell and cancer cell phase image generated by the supervised model and the diffusion model trained using all and limited data. The top of the result is the entire phase image, and the bottom is the cropped image of the cell with the red box enlarged. The results are separated into comparison results that trained only red blood cells or cancer cells and comparison results that simultaneously trained red blood cells or cancer cells. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 5
PSNR and SSIM values for phase images generated by models simultaneously trained on red blood cells and cancer cells from 200 test images (left: PSNR, right: SSIM).

Model	Red blood cells	Cancer cells
Supervised model	34.68/0.8996	35.32/0.9067
Supervised model with 50 labels	30.23/0.8277	30.39/0.8136
Diffusion model	34.26/0.8359	35.27/0.8591
Diffusion model with 50 labels	33.04/0.8084	33.08/0.8305

supervised model was significantly lowered. Nevertheless, the supervised model performs well with red blood cells, likely because red blood cells have more consistent morphology compared to cancer cells. Because the morphology of red blood cells is similar among cells, it is easy to extract cell characteristics even when using a small amount of data. Conversely, the varied morphology among cancer cells presents challenges in feature extraction and training with limited data. The unsupervised diffusion model, optimized to train on the general characteristics of the domain, demonstrates superior performance even with fewer labels compared to the supervised model.

3.4. Noise schedule parameters sensitivity analysis

The performance of diffusion models is critically dependent on the choice of hyperparameters, particularly the noise schedule parameters that govern the forward diffusion process. To ensure reproducibility and understand the robustness of our approach, we conducted a comprehensive sensitivity analysis on key hyperparameters: the number of timesteps (T) and the noise variance bounds ($\overline{\beta}_{\min}$ and $\overline{\beta}_{\max}$) using models simultaneously trained on both red blood cells and cancer cells. This analysis is essential for validating the stability of our model and providing guidance for parameter selection in different experimental conditions. The hyperparameters used in our main experiments ($T = 4$, $\overline{\beta}_{\min} = 0.1$, $\overline{\beta}_{\max} = 20$) were adopted from the denoising diffusion GAN framework proposed by Özbey et al., which has demonstrated effectiveness in medical imaging applications. However, given the unique characteristics of holographic phase reconstruction, it is crucial to evaluate how variations in these parameters affect reconstruction quality. We systematically varied timesteps ($T = 2, 6, 8$) and noise variance bounds across three configurations to assess their impact on phase reconstruction performance.

The experimental results presented in Table 6 demonstrate clear relationships between hyperparameter choices and reconstruction quality across different cell types. For timestep sensitivity, we observed progressive improvements in both SSIM and PSNR metrics as the number of timesteps increased. Cancer cells showed SSIM values improving from 0.814 ($T = 2$) to 0.836 ($T = 6$) and 0.849 ($T = 8$), with corresponding PSNR increases from 34.32 to 34.58 and 34.98 dB respectively. Red blood cells exhibited similar trends with SSIM values of 0.788, 0.814, and 0.829, and PSNR values of 34.32, 34.42, and 34.93 dB. Notably, our main configuration ($T = 4, \overline{\beta}_{\min} = 0.1, \overline{\beta}_{\max} = 20$) achieved the highest performance with RBC: 34.26 dB/0.8359 and Cancer: 35.27 dB/0.8591, surpassing all tested variations. Among the noise variance configurations, moderate settings ($\overline{\beta}_{\min} = 0.05$, $\overline{\beta}_{\max} = 10$) showed competitive performance with SSIM values of 0.843 (cancer) and 0.841 (red blood cells), while excessive noise variance led to degradation. This validates our parameter selection and demonstrates that balanced noise scheduling is crucial for optimal phase reconstruction.

Table 6
Hyperparameter sensitivity analysis results for diffusion model trained on both red blood cells and cancer cells from 200 test images (left: PSNR, right: SSIM).

Model	Red blood cells	Cancer cells
Main ($T = 4, \overline{\beta} = 0.1/20$)	34.26/0.8359	35.27/0.8591
$T = 2, \overline{\beta} = 0.1/20$	33.20/0.7882	34.32/0.8144
$T = 6, \overline{\beta} = 0.1/20$	33.78/0.8148	34.58/0.8358
$T = 8, \overline{\beta} = 0.1/20$	34.31/0.8294	34.98/0.8485
$T = 4, \overline{\beta} = 0.05/10$	33.20/0.8410	34.29/0.8431
$T = 4, \overline{\beta} = 0.2/40$	33.71/0.7959	34.58/0.8225
$T = 4, \overline{\beta} = 0.1/50$	34.11/0.8239	34.97/0.8545

3.5. Comparison with iterative phase reconstruction methods

To evaluate the feasibility of single-shot phase reconstruction from in-line holograms, we conducted comparative experiments with iterative phase reconstruction methods. Two distinct approaches were tested: conventional numerical iterative algorithm (Latychevskaia, 2019) and deep learning-based iterative method (Deep DIH) (Li et al., 2020). The numerical approach utilized established phase retrieval techniques such as the Gerchberg-Saxton (GS) algorithm, which employs constraint-based optimization through alternating projections between the hologram and object planes. The deep learning-based method leveraged neural network architectures to learn the mapping between holographic measurements and phase distributions. All experimental comparisons were standardized using 1000 iterations to ensure fair evaluation across different algorithmic approaches.

The comparative results, demonstrated in Fig. 13, reveal significant differences in reconstruction quality between methods. The GS algorithm shows partial success in recovering red blood cell morphology but exhibits substantial background artifacts. Deep DIH provides better background suppression but fails to accurately reconstruct cellular structures. The proposed diffusion model achieves superior reconstruction fidelity, closely matching ground truth for both cellular detail and background clarity. Quantitative evaluation in Table 7 confirms these observations, with the diffusion model achieving 34.26 dB PSNR/0.8359 SSIM for red blood cells and 35.27 dB/0.8591 for cancer cells, substantially outperforming GS (RBC 17.15 dB/0.6525, Cancer 25.48 dB/0.7461) and Deep DIH (RBC 21.93 dB/0.7333, Cancer 30.37 dB/0.8220). The superior performance of Deep DIH over GS stems from better background reconstruction capability.

These reconstruction challenges arise from fundamental limitations in single-shot in-line holography. As established by Latychevskaia and Fink, twin image interference creates severe problems for complex biological specimens with varying optical densities and irregular morphologies. Biological samples present additional complications including weak phase contrast, irregular boundaries, and heterogeneous refractive indices that violate sparse object assumptions underlying conventional algorithms. Such factors make convergence challenging, rendering traditional iterative methods unreliable for biological specimens. Additionally, the proposed method demonstrates exceptional computational efficiency, requiring only 0.271 s per one image compared to GS (475.80 s) and Deep DIH (286.09 s). This advantage stems from direct inference through pre-trained networks versus iterative optimization processes, making conventional approaches impractical for real-time applications.

3.6. Robustness evaluation across different propagation distances

One critical limitation of diffusion-based holographic reconstruction models is their potential sensitivity to variations in optical parameters, particularly propagation distance, which directly affects hologram formation and phase reconstruction quality. To evaluate the robustness of our approach under varying imaging conditions, we conducted a comprehensive analysis comparing models trained on two distinct datasets: one with randomized propagation distances and another with fixed in-focus distances. For red blood cells imaged at 40 × magnification, training data was acquired with randomly adjusted distances within ±1.8 μm from the in-focus position, while cancer cells at 20 × magnification utilized a ±3.0 μm range. Test datasets were

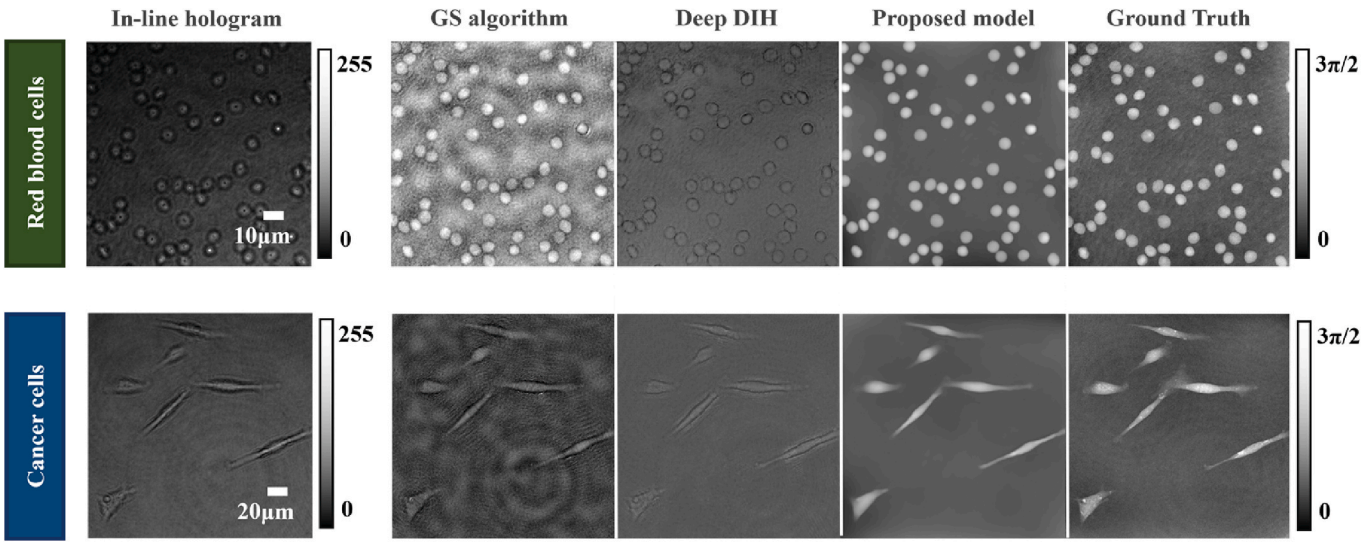


Fig. 13. Comparative reconstruction results from single-shot in-line holograms of biological samples. The top row shows red blood cells, and the bottom row shows cancer cells. From left to right: original in-line hologram, GS algorithm result, Deep DIH result, proposed diffusion model result, and ground truth obtained from off-axis holography. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 7
Quantitative comparison of reconstruction methods showing PSNR/SSIM values for red blood cells and cancer cells, along with inference times.

Model	Red blood cells	Cancer cells	Inference time
GS-based algorithm	17.15/0.6525	25.48/0.7461	475.80 s
Deep DIH	21.93/0.7333	30.37/0.8220	286.09 s
Diffusion model	34.26/0.8359	35.27/0.8591	0.271 s

systematically generated at 0.3 μm intervals (RBCs) and 0.5 μm intervals (cancer cells) across their respective ranges, with five images acquired at each distance position. This experimental design allows us to assess how training data diversity in propagation distance affects model generalization and reconstruction fidelity under defocused conditions, addressing practical scenarios where precise focus control may be challenging.

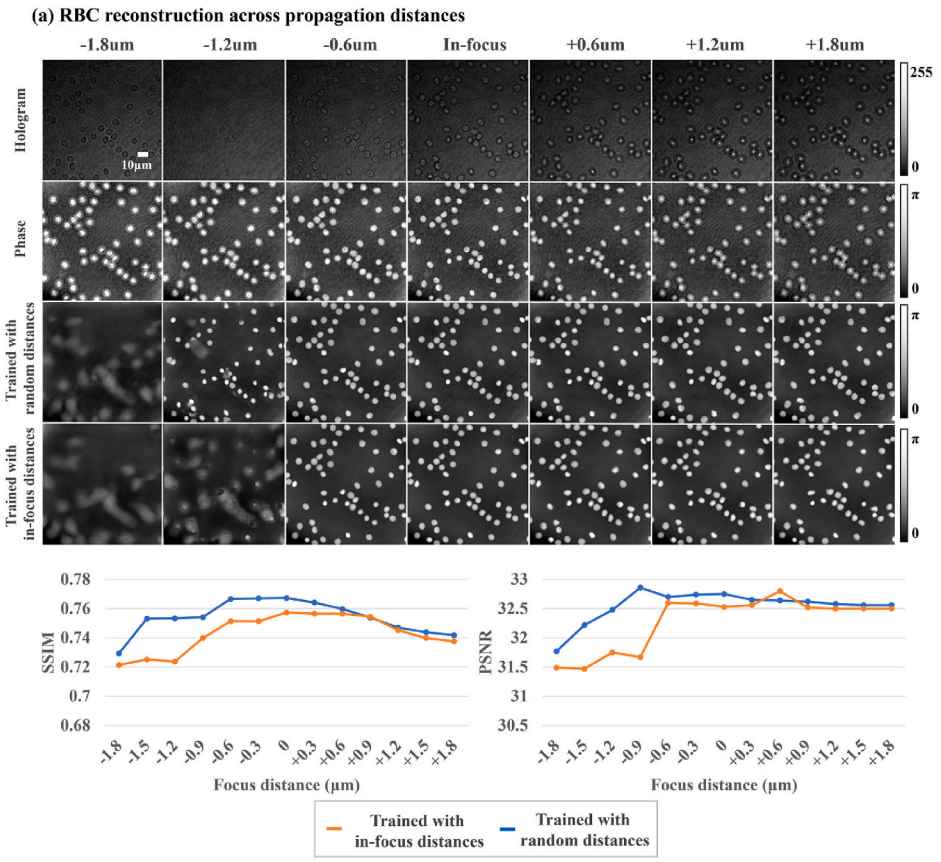
The results for red blood cells, as shown in Fig. 14a, reveal significant sensitivity to negative propagation distance variations due to the high magnification (40 ×) used for imaging. Both models demonstrate acceptable reconstruction quality within the positive distance range, where the in-line holograms remain relatively similar to the in-focus condition. However, performance degrades substantially for negative distances beyond −0.9 μm, as evidenced by sharp drops in both SSIM and PSNR metrics. The model trained with random distances shows robustness, maintaining recognizable cellular morphology even at −1.2 μm, while the in-focus trained model loses cellular features beyond this threshold. This performance difference highlights the importance of training data diversity in achieving robust reconstruction across varying optical conditions. The asymmetric performance between positive and negative distances suggests that the in-focus hologram characteristics are more similar to slightly positive-defocused conditions, explaining the better reconstruction quality in the positive direction. These findings emphasize the need for careful consideration of propagation distance variations during model training for high-magnification applications.

Cancer cell reconstruction demonstrates significantly improved robustness across different propagation distances, as illustrated in Fig. 14b, primarily due to the lower magnification (20 ×) that reduces sensitivity to distance variations. Both training approaches maintain cellular morphology visibility across the entire tested range, with quantitative metrics showing less degradation compared to red blood cells. For models trained with random distances, negative distances still present challenges, but the performance reduction is substantially less severe, and cellular structures remain clearly distinguishable even at extreme positions. The comparison between random distance and in-focus training reveals consistent but gradual performance differences, with the random distance approach showing slightly superior SSIM and PSNR values across most tested positions. Notably, the performance gap between the two training strategies is considerably smaller for cancer cells than for red blood cells, indicating that lower magnification imaging provides inherent robustness to distance variations. These results demonstrate that our diffusion model can effectively reconstruct cancer cell phase images across a practical range of propagation distances, making it suitable for applications where precise focus control may be challenging or where rapid imaging protocols are required.

3.7. Generalization evaluation on different cancer cell morphologies

To evaluate the generalization capability of our diffusion model across diverse cellular structures, we conducted experiments using two distinct training configurations with varying levels of morphological diversity. The single-cancer configuration utilized our standard dataset comprising red blood cells and ovarian cancer cells (SKOV3), while the multi-cancer approach incorporated red blood cells alongside four different cancer cell lines representing various origins: ovarian (SKOV3), liver (SNU475), bladder (T24), and lung (NCI-H1299). As illustrated in Fig. 15, these cancer cell lines exhibit significantly different morphological characteristics compared to the original ovarian cancer cells, including variations in cell size, shape complexity, and internal structures. The liver cancer cells display larger, more circular morphologies, while bladder cancer cells show elongated structures, and lung cancer cells present irregular, highly variable shapes. This morphological diversity provides a robust testing ground for assessing the model's ability to generalize beyond its training data and reconstruct phase images from previously unseen cellular architectures, which is crucial for practical biomedical applications.

When examining reconstruction quality across different cancer cell morphologies, we observed consistent patterns in performance characteristics. As shown in Fig. 15, while reconstructed phase images maintain good overall fidelity to ground truth measurements, these morphologically diverse cell types exhibit slight edge blurring at cellular boundaries, particularly in regions with complex structural features. Despite this visual edge softening, quantitative metrics demonstrate high reconstruction quality.



(caption on next page)

Fig. 14. Robustness evaluation across different propagation distances. Two training regimes are compared: a model trained only at the in-focus distance and a model trained with randomized propagation distances. For each dataset, the image montage shows representative results across defocus and the line plots report SSIM/PSNR versus focus distance. (a) RBC results (40 ×). Training distances were randomized within ±1.8 μm; testing was performed every 0.3 μm with five images per position. Columns span −1.8 → +1.8 μm; rows show (from top) input hologram, reference phase, reconstruction from the in-focus–trained model, and reconstruction from the random-distance–trained model. (b) Cancer-cell results (20 ×). Training distances were randomized within ±3.0 μm; testing used 0.5 μm steps with five images per position. The same row order is used as in (a), with columns spanning −3.0 → +3.0 μm.

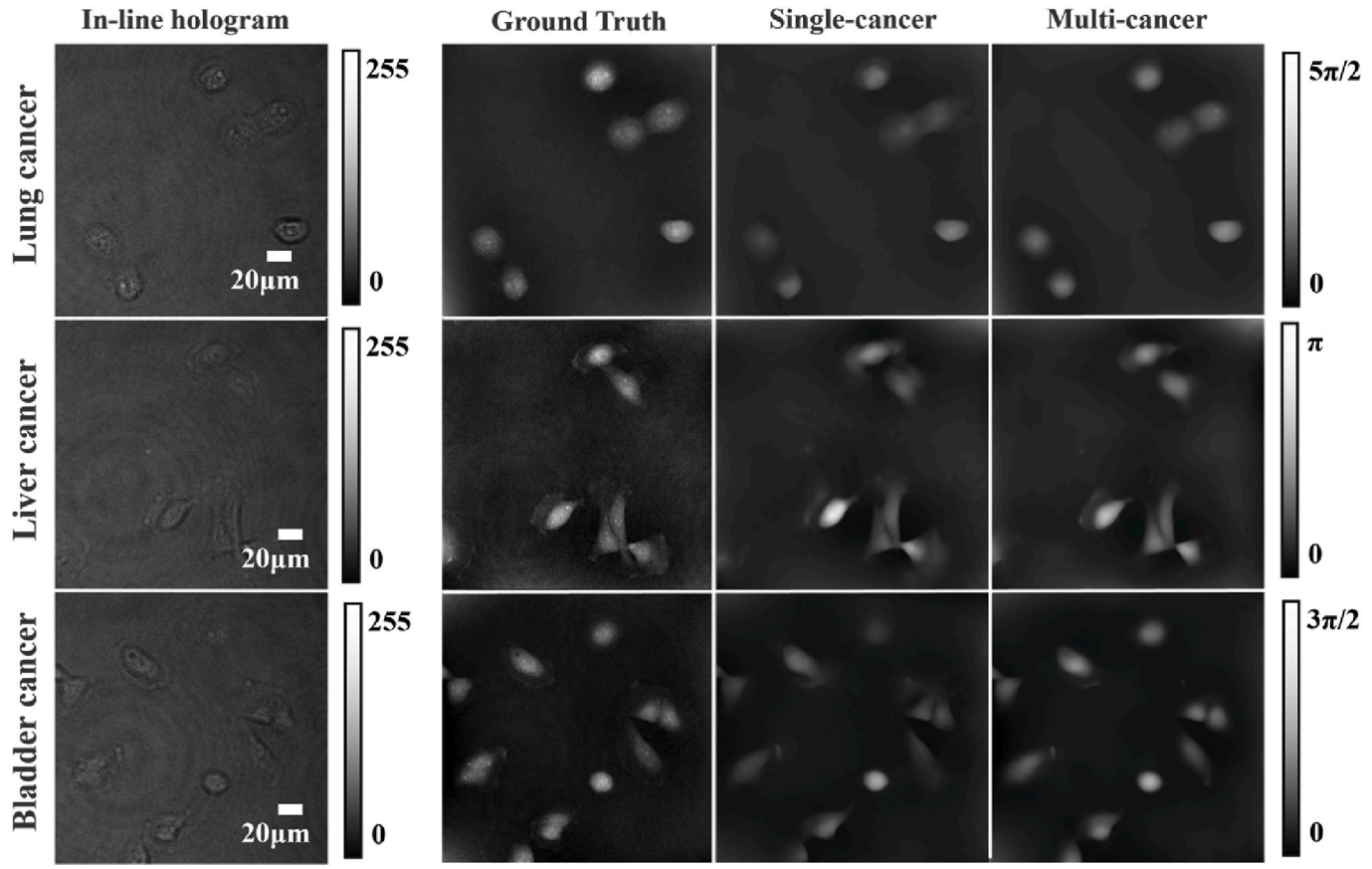


Fig. 15. Generalization of diffusion-based phase reconstruction across diverse cancer cell morphologies. To assess the model's ability to generalize beyond its training data, we compared two training configurations: a single-cancer setup trained on RBCs and ovarian cancer cells, and a multi-cancer setup trained on RBCs plus four cancer cell lines (ovarian, liver, bladder, and lung). Representative examples are shown for three previously unseen cancer types: lung, liver, and bladder. Columns display the in-line holograms (left), ground-truth off-axis phase reconstructions (middle-left), single-cancer trained outputs (middle-right), and multi-cancer trained outputs (right).

As summarized in Table 8, the comparative analysis reveals interesting trade-offs in model performance when training data diversity is increased through multi-cancer training. Models trained with the multi-cancer approach demonstrated improved reconstruction performance for cancer cells, with enhanced SSIM and PSNR metrics compared to the single-cancer training configuration. However, this improvement came at the cost of slightly reduced performance for red blood cell reconstruction. Notably, the single-cancer trained model still achieved reasonable reconstruction quality for the three previously unseen cancer cell lines (liver, bladder and lung), despite never encountering these morphologies during training. This demonstrates the diffusion model's inherent ability to learn generalizable reconstruction principles rather than simply memorizing specific cellular features. The successful reconstruction of morphologically distinct cancer cells suggests that the model has effectively learned the underlying physics of phase image formation and holographic reconstruction processes.

Table 8
Reconstruction performance comparison between single-cancer and multi-cancer training configurations across red blood cells and four cancer cell lines from 200 test images (left: PSNR, right: SSIM).

Training data	Red blood cells	Ovarian cancer	Lung cancer	Liver cancer	Bladder cancer
Single-cancer	34.26/0.8359	35.27/0.8591	35.70/0.8753	35.90/0.8724	34.45/0.8567
Multi-cancer	33.88/0.7867	35.63/0.8766	36.20/0.8787	36.62/0.8767	35.21/0.8625

3.8. Performance analysis with wrapped phase data

Phase wrapping is a fundamental challenge in quantitative phase imaging, where phase values are constrained to the range [-π, π], creating discontinuities at boundaries where the actual phase exceeds these limits. To assess our diffusion model's capability in handling different phase representations, we conducted comparative experiments using wrapped and unwrapped phase images as training targets. Wrapped phase images contain artificial discontinuities where phase values jump from +π to -π (or vice versa), while unwrapped phase images maintain continuous phase distributions that better represent the true optical path differences through biological specimens. This

comparison is particularly relevant for practical applications where phase unwrapping algorithms may be unavailable or computationally expensive, potentially making wrapped phase reconstruction an attractive alternative. The experimental setup utilized the same in-line hologram inputs but different ground truth phase representations, allowing direct assessment of how phase representation affects reconstruction quality. Understanding these differences is crucial for determining optimal training strategies and evaluating the model's robustness across different phase imaging scenarios commonly encountered in digital holographic microscopy applications.

The experimental results demonstrate significant performance differences between wrapped and unwrapped phase training approaches, as illustrated in Fig. 16. Models trained with wrapped phase data struggled to accurately reconstruct continuous phase transitions, particularly in regions where phase wrapping occurred. While reconstruction quality remained reasonable in areas without phase discontinuities, the model failed to properly identify and handle phase transition boundaries, resulting in artifacts and discontinuous phase profiles. In contrast, models trained with unwrapped phase data successfully reconstructed continuous phase distributions, maintaining smooth transitions across the entire cellular structure. The quantitative analysis reveals that while PSNR and SSIM metrics appear relatively similar (wrapped: RBC 33.23/0.7764, Cancer 34.36/0.8225; unwrapped: RBC 34.26/0.8359, Cancer 35.27/0.8591), this similarity is primarily due to most adherent cells exhibiting limited phase wrapping (Table 9). However, the unwrapped training consistently outperformed wrapped

Table 9
Reconstruction performance comparison for diffusion models trained with wrapped versus unwrapped phase images on red blood cells and cancer cells from 200 test images (left: PSNR, right: SSIM).

Training data	Red blood cells	Cancer cells
Wrapped phase	33.23/0.7764	34.36/0.8225
Unwrapped phase	34.26/0.8359	35.27/0.8591

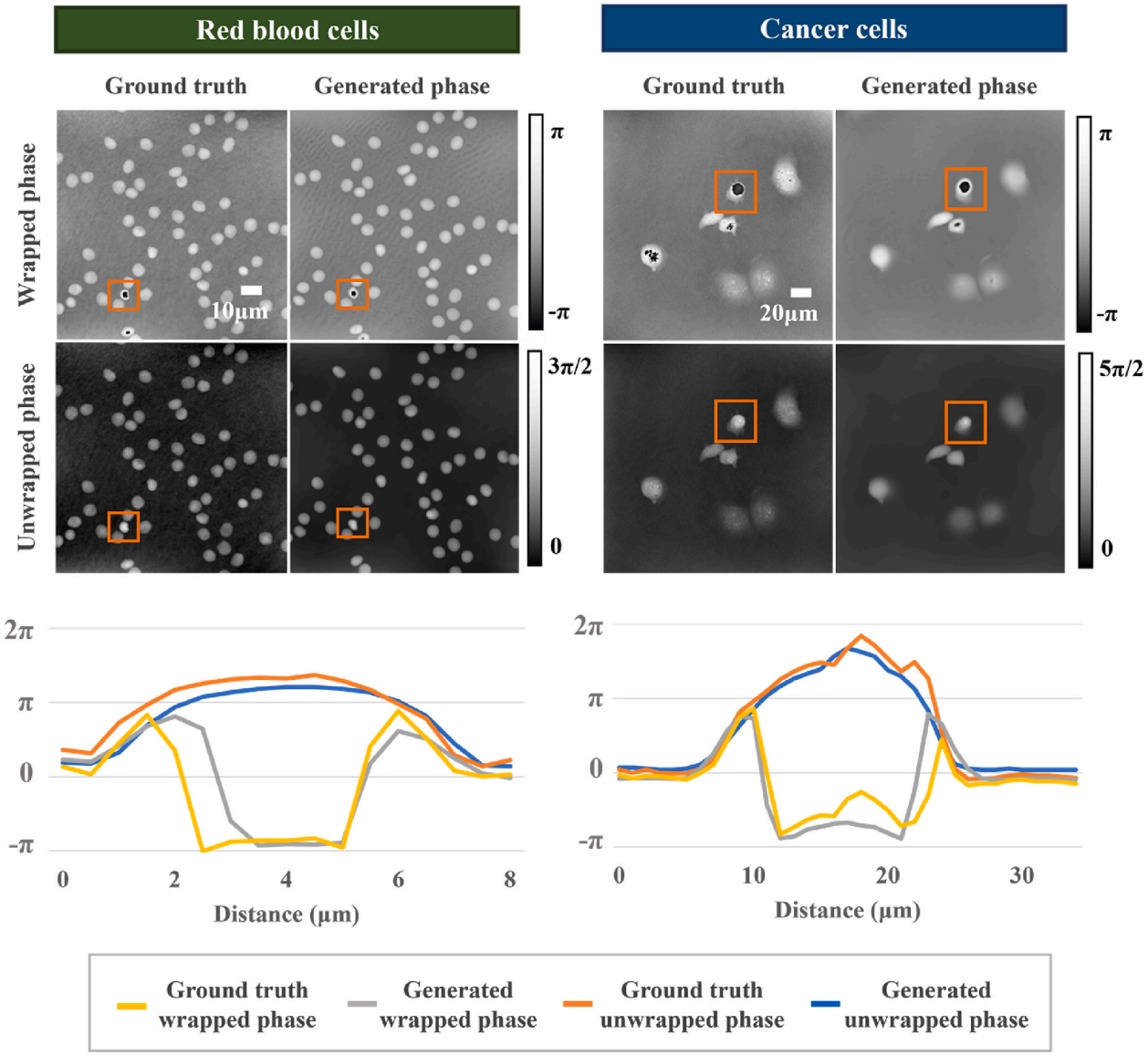


Fig. 16. Comparison of reconstruction performance using wrapped versus unwrapped phase targets. Left panels show results for red blood cells (RBCs) and right panels show results for cancer cells. For each dataset, the top row corresponds to models trained with wrapped phase images (ground truth vs. generated), and the second row corresponds to models trained with unwrapped phase images. Orange boxes indicate representative cells selected for line profile analysis. The bottom graphs plot the phase distributions across the highlighted cells, comparing ground-truth wrapped, generated wrapped, ground-truth unwrapped, and generated unwrapped results. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

training across both cell types. The line profile analysis clearly demonstrates the superior ability of unwrapped training to maintain phase continuity, making it the preferred approach for quantitative phase reconstruction applications requiring accurate optical path difference measurements.

4. Discussion

4.1. Comparison with alternative compact phase imaging technologies

Several compact phase imaging technologies have emerged as alternatives to traditional off-axis holography, each offering distinct advantages and limitations. Quadriwave Lateral Shearing Interferometry (QLSI) provides single-shot quantitative phase imaging using a specialized diffraction grating and is commercially available with excellent stability (Bon et al., 2009). However, QLSI suffers from reduced spatial resolution (typically 4 $\times$ lower than the sensor resolution) and requires expensive specialized hardware. Coded Wavefront Sensing achieves ultra-high spatial resolution through spatial light modulators and computational algorithms, but faces significant limitations in acquisition speed due to the need for multiple sequential measurements (8–20 acquisitions) and computationally intensive iterative reconstruction algorithms, making it impractical for real-time applications (Kazim et al., 2025; Wang et al., 2017). Self-reference digital holography with LED illumination offers speckle-free imaging using everyday light sources, but requires complex optical setups with polarization-sensitive components and precise alignment procedures (Tahara, 2024).

In contrast, our unsupervised diffusion model approach offers several distinct advantages: Cost-effectiveness - utilizing standard Gabor holography setups without specialized hardware, Real-time capability - single-shot acquisition with rapid inference once trained, Scalability - software-based improvements through advanced AI algorithms, and Data efficiency - robust performance even with limited training data through unsupervised learning. While our method may not achieve the absolute highest spatial resolution of coded wavefront sensing, it provides an optimal balance of performance, practicality, and accessibility for quantitative phase imaging applications.

4.2. Spatial resolution considerations and trade-offs

In our study, an apparent gap exists between the theoretical resolution defined by the optical system and the practical resolution observed in the reconstructed phase images. The theoretical transverse resolution of both off-axis and in-line configurations is determined by the diffraction limit of the microscope objective (MO), expressed as $\delta \approx 0.61\lambda/\mathrm{NA}$, and is further influenced by the pixel size of the image sensor. In practice, however, the effective resolution differs depending on the acquisition geometry and reconstruction strategy. Off-axis DHM benefits from the Fourier-domain separation of real, twin, and zero-order terms, enabling it to approach its diffraction-limited resolution more closely. In contrast, in-line (Gabor) holography, while theoretically capable of a larger space–bandwidth product, suffers from twin-image overlap, background contributions, and speckle noise that degrade fine spatial details. Moreover, our AI-based reconstruction framework is trained using phase images from off-axis DHM as ground truth, which constrains the reconstructed outputs to the resolution ceiling imposed by the NA and magnification of the off-axis system.

Despite these limitations, our method achieves effective resolution sufficient for cellular morphology analysis and quantitative phase measurements, as demonstrated in red blood cell and cancer cell reconstructions. Importantly, this trade-off reflects a deliberate design choice: unlike hardware-intensive methods such as coded wavefront sensing, which can achieve ultra-high resolution through multiple measurements, our single-shot approach prioritizes simplicity and temporal resolution, making it particularly suitable for dynamic imaging scenarios where speed and cost-efficiency are critical. Future work will aim to bridge the resolution gap by incorporating deep-learning-based super-resolution strategies, physics-informed priors, and advanced denoising architectures with multi-scale training, thereby enhancing spatial fidelity while maintaining computational efficiency and robustness to noise.

4.3. Training versus inference: understanding the single-shot nature of our approach

A critical aspect of our methodology that requires clear explanation is the fundamental distinction between the training phase and the inference phase, which directly addresses why our approach can be considered a true single-shot system despite requiring off-axis data during training. During the training phase, our unsupervised diffusion model learns the complex mapping relationship between in-line holographic patterns and their corresponding quantitative phase distributions using unpaired datasets. This training process requires access to both in-line holograms (captured in cost-effective Gabor mode) and high-quality reference phase images (obtained from off-axis digital holographic reconstruction). The off-axis measurements serve exclusively as reliable ground truth targets that enable the model to understand the relationship between simple holographic interference patterns and the underlying cellular phase information. This training phase represents a one-time investment in computational learning that establishes the model's reconstruction capabilities.

The inference phase, however, operates fundamentally differently and represents the true operational mode of our system. Once training is completed, the model requires only a single in-line hologram as input to generate high-quality quantitative phase reconstructions. No off-axis measurements, complex optical alignments, or additional reference data are needed during actual use. This single-input, single-output operation definitively qualifies our approach as a single-shot system for practical deployment. The key advantage of this paradigm is that the computational complexity and data requirements are front-loaded into the training phase, while the operational phase remains remarkably simple and cost-effective. A single trained model can be deployed across multiple simple in-line holographic setups, transforming basic Gabor-mode systems into quantitative phase imaging platforms without requiring expensive off-axis infrastructure at each deployment location. This represents a fundamental shift from traditional approaches where each measurement location must be equipped with complex optical systems.

4.4. Advantages and limitations of the proposed approach

Our unsupervised diffusion model demonstrates several compelling advantages over existing phase reconstruction methodologies. The approach exhibits superior generalization capabilities across diverse cellular morphologies, as demonstrated by successful reconstruction of multiple cancer cell lines without requiring cell-specific training. The model's robustness under varying propagation distances and its ability to maintain performance with limited training data highlight its practical applicability in resource-constrained environments. Furthermore, the unsupervised learning framework eliminates the need for extensive labeled datasets, significantly reducing data preparation overhead compared to supervised approaches. The iterative denoising process inherent to diffusion models provides enhanced noise resilience and reconstruction stability compared to single-pass GAN-based methods. However, limitations include increased computational requirements during training and inference compared to traditional analytical methods, sensitivity to extreme defocus conditions particularly at high magnifications, and the current restriction to learned optical parameter ranges. The model's performance degradation with wrapped phase data also indicates limitations in handling certain phase imaging scenarios. Despite these constraints, the balance of reconstruction quality, practical deployment feasibility, and cost-effectiveness positions our

approach as a viable solution for quantitative phase imaging applications.

4.5. Future directions and potential improvements

Several promising avenues exist for enhancing the capabilities and applicability of our diffusion-based phase reconstruction framework. Integration of physics-informed constraints into the diffusion process could improve reconstruction accuracy by incorporating known optical propagation principles, potentially addressing current limitations with extreme defocus conditions and wrapped phase handling. Multi-scale training strategies using different magnifications and imaging parameters could enhance generalization across diverse optical configurations, reducing the need for system-specific retraining. Advanced attention mechanisms and transformer architectures could improve the model's ability to capture long-range spatial dependencies in phase distributions, potentially enhancing resolution capabilities. Real-time optimization through model compression techniques, knowledge distillation, and specialized hardware acceleration could reduce inference times to enable true real-time applications. Additionally, extending the framework to handle 3D phase reconstruction and multi-wavelength imaging could broaden its applicability to advanced holographic microscopy techniques. Incorporation of uncertainty quantification mechanisms would provide confidence measures for reconstruction quality, enabling automated quality control in clinical applications. Finally, development of domain adaptation strategies could facilitate rapid deployment across different biological specimens and imaging systems without requiring extensive retraining, making the technology more accessible for diverse research and clinical environments.

5. Conclusion

This study introduces the first unsupervised diffusion model for automated phase image reconstruction from single-shot in-line holograms, addressing the fundamental twin image problem that has long limited Gabor holography applications. Our computational framework successfully eliminates the need for expensive off-axis optical setups while maintaining reconstruction fidelity comparable to traditional complex systems. The key breakthrough lies in enabling effective training on unpaired hologram-phase datasets through integrated cycle-consistency and denoising modules. This approach demonstrates superior performance compared to existing unsupervised methods (CycleGAN, UNIT) and maintains exceptional reconstruction quality even with severely limited training data—a critical advantage over supervised approaches that suffer significant degradation under data-constrained conditions. Particularly notable is the model's remarkable generalization capability across diverse biological specimens. The framework successfully reconstructs phase images from various cancer cell morphologies not encountered during training and maintains robustness across different propagation distances, demonstrating its practical applicability in real-world imaging scenarios where precise optical control may be challenging. Our approach represents a paradigm shift from hardware-based to computational solutions for quantitative phase imaging. By transforming low-cost, simple optical measurements into high-quality phase reconstructions through advanced AI algorithms, this work democratizes access to quantitative phase microscopy for resource-constrained environments and real-time applications. This computational breakthrough opens new possibilities for compact digital holography systems in biomedical research, clinical diagnostics, and field applications, where cost-effectiveness, portability, and ease of operation are paramount while maintaining the quantitative accuracy required for meaningful biological analysis.

CRediT authorship contribution statement

Seonghwan Park: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Jaewoo Park:** Visualization, Investigation. **Youhyun Kim:** Writing – review & editing, Validation. **Inkyu Moon:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **Bahram Javidi:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. RS-2023-00253613) and by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. RS-2025-25420118).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engappai.2025.112970.

Data availability

Data will be made available on request.

References

Ahamadzadeh, E., Jaferzadeh, K., Park, S., Son, S., Moon, I., 2022. Automated analysis of human cardiomyocytes dynamics with holographic image-based tracking for cardiotoxicity screening. Biosens. Bioelectron. 195.

Barbastathis, G., Ozcan, A., Situ, G., 2019. On the use of deep learning for computational imaging. Optica 6, 921–943.

Bhabhor, G.D., Patel, C., Chhillar, N., Anand, A., Lad, K.N., 2024. Geometrical characterization of healthy red blood cells using digital holographic microscopy and parametric shape models for biophysical studies and diagnostic applications. J. Phys. Appl. Phys. 57, 355401.

Bon, P., Maucort, G., Wattellier, B., Monneret, S., 2009. Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells. Opt. Express 17, 13080–13094.

Chen, W., 2019. Single-shot in-line holographic authentication using phase and amplitude modulation. Opt Laser. Eng. 121, 473–478.

Colomb, T., Cuche, E., Charrière, F., Kühn, J., Aspert, N., Montfort, F., Marquet, P., Depeursinge, C., 2006. Automatic procedure for aberration compensation in digital holographic microscopy and applications to specimen shape compensation. Appl. Opt. 45, 851–863.

Cuche, E., Marquet, P., Depeursinge, C., 1999. Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of fresnel off-axis holograms. Appl. Opt. 38, 6994–7001.

Dardikman-Yoffe, G., Roitshtain, D., Mirsky, S.K., Turko, N.A., Habaza, M., Shaked, N.T., 2020. PhUn-Net: Ready-to-use neural network for unwrapping quantitative phase images of biological cells. Biomed. Opt. Express 11, 1107–1121.

Gabor, D., 1948. A new microscopic principle. Nature 161, 777–778.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Adv Neur In 27, 2672–2680.

He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), pp. 770–778.

Herráez, M.A., Burton, D.R., Lalor, M.J., Gdeisat, M.A., 2002. Fast two-dimensional phase-unwrapping algorithm based on sorting by reliability following a noncontinuous path. Appl. Opt. 41, 7437–7444.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Adv. Neural Inf. Process. Syst. 33, 6840–6851.

Jaferzadeh, K., Fevens, T., 2022. HoloPhaseNet: fully automated deep-learning-based hologram reconstruction using a conditional generative adversarial model. Biomed. Opt. Express 13, 4032–4046.

Jaferzadeh, K., Moon, I., 2016. Human red blood cell recognition enhancement with three-dimensional morphological features obtained by digital holographic imaging. J. Biomed. Opt. 21.

Jaferzadeh, K., Moon, I., Bardyn, M., Prudent, M., Tissot, J.D., Rappaz, B., Javidi, B., Turcatti, G., Marquet, P., 2018. Quantification of stored red blood cell fluctuations by time-lapse holographic cell imaging. Biomed. Opt. Express 9, 4714–4729.
Javidi, B., Carnicer, A., Anand, A., Barbastathis, G., Chen, W., Ferraro, P., Goodman, J. W., Horisaki, R., Khare, K., Kujawinska, M., Leitgeb, R.A., Marquet, P., Nomura, T., Ozcan, A., Park, Y., Pedrini, G., Picart, P., Rosen, J., Saavedra, G., Shaked, N.T., Stern, A., Tajahuerce, E., Tian, L., Wetzstein, G., Yamaguchi, M., 2021. Roadmap on digital holography [Invited]. Opt. Express 29, 35078–35118.
Kazim, S.M., Strasser, F., Kvåle Løvmo, M., Nehrych, A., Moser, S., Ziemczonok, M., Heidrich, W., Ihrke, I., Ritsch-Marte, M., 2025. Coded wavefront sensing for video-rate quantitative phase imaging and tomography: validation with digital holographic microscopy. Opt. Express 33, 25198–25209.
Kemper, B., von Bally, G., 2008. Digital holographic microscopy for live cell applications and technical inspection. Appl. Opt. 47, A52–A61.
Kim, E., Park, S., Hwang, S., Moon, I., Javidi, B., 2022. Deep learning-based phenotypic assessment of red cell storage lesions for safe transfusions. IEEE J Biomed Health Inform 26, 1318–1328.
Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv: 1312.6114.
Koren, G., Polack, F., Joyeux, D., 1993. Iterative algorithms for twin-image elimination in in-Line holography using finite-support constraints. J. Opt. Soc. Am. A 10, 423–433.
Lam, V.K., Nguyen, T.C., Chung, B.M., Nehmetallah, G., Raub, C.B., 2018. Quantitative assessment of cancer cell morphology and motility using telecentric digital holographic microscopy and machine learning. Cytometry A 93, 334–345.
Latychevskaia, T., 2019. Iterative phase retrieval for digital holography: tutorial. J. Opt. Soc. Am. A 36, D31–D40.
Latychevskaia, T., Fink, H.-W., 2007. Solution to the twin image problem in holography. Phys. Rev. Lett. 98, 233901.
Li, H., Chen, X., Chi, Z., Mann, C., Razi, A., 2020. Deep DIH: single-shot digital in-line holography reconstruction by deep learning. IEEE Access 8, 202648–202659.
Liu, M.-Y., Breuel, T., Kautz, J., 2017. Unsupervised image-to-image translation networks. Adv. Neural Inf. Process. Syst. 30.
Manisha, Mandal, A.C., Rathor, M., Zalevsky, Z., Singh, R.K., 2023. Randomness assisted in-line holography with deep learning. Sci. Rep. 13, 10986.
Moon, I., 2022. Artificial Intelligence in Digital Holographic Imaging: Technical Basis and Biomedical Applications. Wiley.
Moon, I., Javidi, B., 2008. 3-D visualization and identification of biological microorganisms using partially temporal incoherent light In-Line computational holographic imaging. Ieee T Med Imaging 27, 1782–1790.
Nakamura, T., Nitta, K., Matoba, O., 2007. Iterative algorithm of phase determination in digital holography for real-time recording of real objects. Appl. Opt. 46, 6849–6853.
O'Connor, T., Anand, A., Andemariam, B., Javidi, B., 2020. Deep learning-based cell identification and disease diagnosis using spatio-temporal cellular dynamics in compact digital holographic microscopy. Biomed. Opt. Express 11, 4491–4508.
O'Connor, T., Shen, J.B., Liang, B.T., Javidi, B., 2021. Digital holographic deep learning of red blood cells for field-portable, rapid COVID-19 screening. Opt. Lett. 46, 2344–2347.
Özbey, M., Dalmaz, O., Dar, S.U.H., Bedel, H.A., Özturk, S., Güngör, A., Cukur, T., 2023. Unsupervised medical image translation with adversarial diffusion models. Ieee T Med Imaging 42, 3524–3539.
Park, S., Kim, Y., Moon, I., 2023. Fast automated quantitative phase reconstruction in digital holography with unsupervised deep learning. Opt Laser. Eng. 167.
Park, S., Kim, Y., Moon, I., 2024. Automated phase reconstruction and super-resolution with deep learning in digital holography. Opt Laser. Technol. 176, 111030.
Pirone, D., Lim, J., Merola, F., Miccio, L., Mugnano, M., Bianco, V., Cimmino, F., Visconte, F., Montella, A., Capasso, M., 2022. Stain-free identification of cell nuclei using tomographic phase microscopy in flow cytometry. Nat. Photonics 16, 851–859.
Pritt, M.D., Shipman, J.S., 1994. Least-squares 2-Dimensional phase unwrapping using ffts. Ieee T Geosci Remote 32, 706–708.
Quiroga, J.A., Bernabeu, E., 1994. Phase-unwrapping algorithm for noisy phase-map processing. Appl. Opt. 33, 6725–6731.
Rappaz, B., Barbul, A., Emery, Y., Korenstein, R., Depeursinge, C., Magistretti, P.J., Marquet, P., 2008. Comparative study of human erythrocytes by digital holographic microscopy, confocal microscopy, and impedance volume analyzer. Cytometry A 73, 895–903.
Rappaz, B., Marquet, P., Cuche, E., Emery, Y., Depeursinge, C., Magistretti, P.J., 2005. Measurement of the integral refractive index and dynamic cell morphometry of living cells with digital holographic microscopy. Opt. Express 13, 9361–9373.
Rivenson, Y., Wu, Y.C., Ozcan, A., 2019. Deep learning in holography and coherent imaging. Light Sci. Appl. 8.
Roitshtain, D., Wolbromsky, L., Bal, E., Greenspan, H., Satterwhite, L.L., Shaked, N.T., 2017. Quantitative phase microscopy spatial signatures of cancer cells. Cytometry A 91, 482–493.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional networks for biomedical image segmentation. Lect. Notes Comput. Sci. 9351, 234–241.
Shaked, N.T., Satterwhite, L.L., Bursac, N., Wax, A., 2010. Whole-cell-analysis of live cardiomyocytes using wide-field interferometric phase microscopy. Biomed. Opt. Express 1, 706–719.
Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B., 2020. Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456.
Tahara, T., 2024. Self-reference digital holography unit adaptable to optical microscope with light-emitting diode. Opt. Express 32, 46090–46102.
Ugele, M., Weniger, M., Stanzel, M., Bassler, M., Krause, S.W., Friedrich, O., Hayden, O., Richter, L., 2018. Label-free high-throughput leukemia detection by holographic microscopy. Adv. Sci. 5, 1800761.
Wang, C., Dun, X., Fu, Q., Heidrich, W., 2017. Ultra-high resolution coded wavefront sensor. Opt. Express 25, 13736–13746.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. Ieee T Image Process 13, 600–612.
Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19.
Xiao, Z., Kreis, K., Vahdat, A., 2021. Tackling the generative learning trilemma with denoising diffusion gans. arXiv preprint arXiv:2112.07804.
Xu, W., Jericho, M.H., Meinertzhagen, I.A., Kreuzer, H.J., 2001. Digital in-line holography for biological applications. Proc. Natl. Acad. Sci. 98, 11301–11305.
Yan, X., Liu, X., Li, J., Zhang, Y., Chang, H., Jing, T., Hu, H., Qu, Q., Wang, X., Jiang, X., 2024. Generating multi-depth 3D holograms using a fully convolutional neural network. Adv. Sci., 2308886
Yi, F., Moon, I., Lee, Y.H., 2015. Three-dimensional counting of morphologically normal human red blood cells via digital holographic microscopy. J. Biomed. Opt. 20.
Yin, D., Gu, Z., Zhang, Y., Gu, F., Nie, S., Ma, J., Yuan, C., 2020. Digital holographic reconstruction based on deep learning framework with unpaired data. IEEE Photon. J. 12, 1–12.
Zhang, Y.H., Noack, M.A., Vagovic, P., Fezzaa, K., Garcia-Moreno, F., Ritschel, T., Villanueva-Perez, P., 2021. PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets. Opt. Express 29, 19593–19604.
Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. Ieee I Conf Comp Vis 2242–2251.