electronics

MDPI

Article

Lensless Digital Holographic Reconstruction Based on the Deep Unfolding Iterative Shrinkage Thresholding Network

Duofang Chen [1,2,3,*], Zijian Guo [1,2], Huidi Guan [1] and Xueli Chen [1,2,3,*]

[1] Center for Biomedical-Photonics and Molecular Imaging, Advanced Diagnostic-Therapy Technology and Equipment Key Laboratory of Higher Education Institutions in Shaanxi Province, Xi'an 710126, China; zijianguo2022@163.com (Z.G.); 22121214053@stu.xidian.edu.cn (H.G.)

[2] Engineering Research Center of Molecular and Neuro Imaging, Ministry of Education & Xi'an Key Laboratory of Intelligent Sensing and Regulation of Trans-Scale Life Information, School of Life Science and Technology, Xidian University, Xi'an 710126, China

[3] Innovation Center for Advanced Medical Imaging and Intelligent Medicine, Guangzhou Institute of Technology, Xidian University, Guangzhou 510555, China

* Correspondence: dfchen@xidian.edu.cn (D.C.); xlchen@xidian.edu.cn (X.C.)

Abstract: Without using any optical lenses, lensless digital holography (LDH) records the hologram of a sample and numerically retrieves the amplitude and phase of the sample from the hologram. Such lensless imaging designs have enabled high-resolution and high-throughput imaging of specimens using compact, portable, and cost-effective devices to potentially address various point-of-care-, global health-, and telemedicine-related challenges. However, in lensless digital holography, the reconstruction results are severely affected by zero-order noise and twin images as only the hologram intensity can be recorded. To mitigate such interference and enhance image quality, extensive efforts have been made. In recent years, deep learning (DL)-based approaches have made significant advancements in the field of LDH reconstruction. It is well known that most deep learning networks are often regarded as black-box models, which poses challenges in terms of interpretability. Here, we present a deep unfolding network, dubbed the ISTAHolo-Net, for LDH reconstruction. The ISTAHolo-Net replaces the traditional iterative update steps with a fixed number of sub-networks and the regularization weights with learnable parameters. Every sub-network consists of two modules, which are the gradient descent module (GDM) and the proximal mapping module (PMM), respectively. The ISTAHolo-Net incorporates the sparsity-constrained inverse problem model into the neural network and hence combines the interpretability of traditional iterative algorithms with the learning capabilities of neural networks. Simulation and real experiments were conducted to verify the effectiveness of the proposed reconstruction method. The performance of the proposed method was compared with the angular spectrum method (ASM), the HRNet, the Y-Net, and the DH-GAN. The results show that the DL-based reconstruction algorithms can effectively reduce the interference of twin images, thereby improving image reconstruction quality, and the proposed ISTAHolo-Net performs best on our dataset.

Keywords: lensless digital holography; deep learning; twin-image elimination

check for updates

Academic Editors: Christos J. Bouras and Stefano Scanzio

Received: 19 February 2025
Revised: 14 April 2025
Accepted: 20 April 2025
Published: 22 April 2025

Citation: Chen, D.; Guo, Z.; Guan, H.; Chen, X. Lensless Digital Holographic Reconstruction Based on the Deep Unfolding Iterative Shrinkage Thresholding Network. *Electronics* **2025**, *14*, 1697. https://doi.org/10.3390/electronics14091697

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Lensless digital holography (LDH) is an innovative imaging technique that captures the three-dimensional (3D) structure of an object without using any traditional optical lenses [1]. By using digital sensors to record holographic interference patterns, LDH enables high-resolution imaging of samples with minimal setup complexity and increased

flexibility. This technology has found applications in various fields, including biomedical imaging, material science, and industrial inspection, due to its ability to provide rich sample information and high spatial resolution [2]. Classical reconstruction methods include the angular spectrum method (ASM), Fresnel transform method, and convolution method [3], which can produce results quickly. However, the reconstruction results are severely affected by zero-order noise and twin images. To mitigate such interference and enhance image quality, extensive efforts have been made [4]. Existing methods involve the multi-height phase retrieval iterative method [5], multi-angle illumination strategy [6], and compressive sensing approach [7]. Despite the advancements in these traditional reconstruction methods, they often exhibit limitations in handling complex interference patterns and noise, which can significantly affect the quality of the reconstructed images [8].

With the rapid development of deep learning (DL) technologies, researchers have applied these advanced methods to the field of lensless digital holographic reconstruction [9–22]. Based on the training strategy and the network model, the DL-based construction algorithms can be categorized into three types: data-driven [9–16], physics-driven [17–20], and hybrid-driven [21–23]. Generally speaking, data-driven reconstruction networks depend on the statistical relation between the measurements and the ground truth data and most of them are designed based on the Unet [24] or the generative adversarial network(GAN) [25] architecture. And large numbers of matched pairs of measurements and labels are needed to train the networks. Physics-driven reconstruction approaches based on untrained networks utilize the optical coherent diffraction theory and no ground truth data are required. But the convergence is slow and typically takes more than several thousand iterations [26]. The hybrid-driven algorithms incorporate the optical diffraction model into the GANs to obtain more robust reconstruction results. Instead of training deep neural networks solely on the basis of given data, the physical forward model is also considered in the neural networks. Table 1 lists the networks mentioned above; in the table, the name or the backbone network, the type, whether it is trained or not, the loss function, and the size of the training dataset are given.

Table 1. Some of the deep learning networks used for LDH reconstruction.

Reference	Net	Type	Trained	Loss Function	Training Dataset Size
Rivenson et al. [9]	DNN	DD	Yes	NM	E:100
Wang et al. [10]	eHoloNet	DD	Yes	MSE	S:9000,E:11623
Ren et al. [11]	HRNet	DD	Yes	MSE	E:8000
Wang et al. [12]	Y-Net	DD	Yes	MSE	E:1332
Yin et al. [13]	CycleGAN	DD	Yes	Cycle-GAN loss	E:2975
Huang et al. [14]	RNN	DD	Yes	GAN loss	NM
Chen et al. [15]	FIN	DD	Yes	MAE	E:600
Kiriy et al. [16]	GAN	DD	Yes	GAN loss	S:15000,E:10000
Wang et al. [17]	PhysenNet	PD	No	MSE	--
Zhang et al. [18]	BlindNet	PD	No	MSE	--
Chen et al. [19]	DHGAN	PD	No	GAN loss	--
An et al. [20]	FNet	PD	No	SSIM+CTV	--
Zhang et al. [21]	PhaseGAN	HD	Yes	GAN loss	E:20000
Tian et al. [22]	GAN	HD	Yes	GAN loss	E:4992
Lee et al. [23]	FMGAN	HD	Yes	GAN loss	E:600

NM: not mentioned, E: real experiment, S: simulation, DD: data-driven, PD: physics-driven, HD: hybrid-driven.

In this work, we present a deep unfolding network which is totally different from the mentioned approaches for LDH reconstruction. The unfolding architecture is believed to be

more interpretable than the traditional DL networks [27] as it replaces the iterative update steps with cascaded networks. As we know, in in-line LDH reconstruction, the object image has sharp edges compared with the twin-image interference, which can be considered as a sparsity prior. Based on the optical diffraction theory, the LDH reconstruction is transformed into an inverse problem with sparsity constraint, which can be solved iteratively by the iterative shrinkage thresholding algorithm (ISTA) [28]. Inspired by the ISTA-Net [27], we propose the ISTAHolo-Net to solve the inverse problem in LDH. The ISTAHolo-Net incorporates the sparsity-constrained physical model into the neural network and decomposes the iterative reconstruction process into a gradient descent module (GDM) and a proximal mapping module (PMM) using convolutional neural networks. One GDM and one PMM constitute a sub-network. And a fixed number of sub-networks are cascaded, each of which is used to replace one update step in the ISTA.

To validate the effectiveness of the network, both simulation and real experiments were conducted, and several metrics were computed and compared. In the simulation experiment, the dataset was constructed based on handwritten digits and the Chinese characters. In the real experiment, the holograms of the USAF resolution chart and EC109 cells were collected to construct the training and testing datasets. The performance of the network under different numbers of sub-networks was studied. Furthermore, we conducted comparison experiments between the ASM [3], HRNet [12], Y-Net [13], DH-GAN, and the proposed ISTAHolo-Net. DH-GAN is based on the architecture in [21] but is changed from the unsupervised model to the supervised one.

The remainder of the paper is organized as follows: Section 2 introduces the proposed network architecture; Section 3 presents the experimental setup and results; Section 4 is the discussion; and Section 5 concludes the paper's work.

2. Materials and Methods

2.1. Principle of LDH

Developed from digital holography, LDH records the holograms directly with an imaging sensor and recovers the object wave field numerically. Illustrated in Figure 1 is a schematic diagram of the LDH imaging process, which mainly involves hologram recording and object image reconstruction.

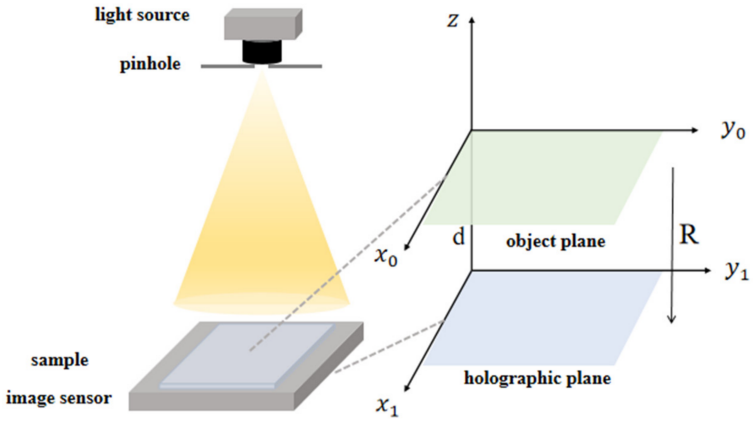


Figure 1. Schematic diagram of lensless digital holographic recording.

In the object plane, the coordinate is denoted as (x_0, y_0), while in the hologram image plane, the coordinate is denoted as (x, y). The object light wave is represented by $O(x, y)$, and the reference light wave is represented by $R(x, y)$. Both of the light waves are complex numbers. When the object light and the reference light have the same frequency and

superimpose on the recording plane, the optical field at any point in space can be described by [3]:

$$F(x,y) = O(x,y) + R(x,y) \tag{1}$$

Therefore, the hologram recorded by the sensor can be expressed by the following formula:

$$\begin{aligned} I(x,y) &= |F(x,y)|^2 \\ &= |O(x,y)|^2 + |R(x,y)|^2 + O(x,y)R^*(x,y) + O^*(x,y)R(x,y) \\ &= |A_o(x,y)|^2 + |A_R(x,y)|^2 + 2A_O(x,y)A_R(x,y)\cos[\varphi_O(x,y) - \varphi_R(x,y)] \end{aligned} \tag{2}$$

where * denotes the complex conjugate. The first two terms, $|O(x,y)|^2$ and $|R(x,y)|^2$, represent the object light intensity and reference light intensity, respectively, which are zero-order images indicating background. $A_O(x,y)$ is the object light amplitude, $A_R(x,y)$ is the reference light amplitude, $\varphi_O(x,y)$ denotes the phase of the object light field, and $\varphi_R(x,y)$ denotes the phase of the reference light. $O(x,y)R^*(x,y)$ stands for the real image, and $O^*(x,y)R(x,y)$ stands for the conjugate image. From Equation (2), it can be seen that in the hologram image, the real image of the object is surrounded by the conjugate image, which blurs the object image and needs to be eliminated. The complex optical field (including amplitude and phase) at the object plane $U(x,y)$ can be reconstructed using ASM from the recorded hologram $H(x,y)$, as follows:

$$U(x,y) = F^{-1}\left[I(f_x,f_y)\cdot H(f_x,f_y)\right] \tag{3}$$

where F^{-1} represents the inverse Fourier transform, $I(f_x,f_y)$ is the Fourier transform of the digital holographic light field distribution, and $H(f_x,f_y)$ is the transfer function of diffraction in the frequency domain.

$$H(f_x,f_y) = \exp\left[jkd\left(1-\lambda^2 f_x^2 - \lambda^2 f_y^2\right)^{1/2}\right] \tag{4}$$

where d is the distance from the object plane to the imaging sensor, and λ is the wavelength of the illumination light source.

2.2. Compressive Sensing (CS)-Based Reconstruction

The holographic reconstruction method based on compressive sensing imposes sparsity constraints on the reconstruction inverse problem, using the sparsity difference between the object image and the twin image to eliminate the twin image [5]. The specific approach to suppress twin images using the CS framework is as follows:

The object's scattered field is defined as:

$$O(x,y) = \iint\limits_{x_i,y_i\in\Omega} \rho(x_i,y_i)h(x-x_i,y-y_i)dx_i dy_i \tag{5}$$

where $\rho(x,y)$ represents the object density, h represents the angular spectrum transfer function, and Ω represents the illumination aperture function. In DLH, the reference light wave is the non-scattered field, which can be assumed to be one without loss of generality and can be removed under plane wave illumination. Therefore, the recorded hologram can be written as:

$$I(x,y) = O^*(x,y) + O(x,y) + |O(x,y)|^2 \tag{6}$$

where $O^*(x,y)$ represents the phase-conjugate field of $O(x,y)$, and $|O(x,y)|^2$ represents the model error, which can be denoted as $e(x,y)$. Thus, the mapping from the diffraction field $O(x,y)$ to the preprocessed hologram can be expressed as:

$$I(x,y) = 2\mathrm{Re}\{O(x,y)\} + e(x,y) \tag{7}$$

We use G to represent the transformation from $O(x,y)$ to $U(x,y)$, as shown in Equation (5); then, the complete mapping from the object density to the preprocessed hologram can be expressed as:

$$I = 2\mathrm{Re}\{H\rho\} + e = G\rho + e \tag{8}$$

The problem of solving for ρ using the known I and the forward transformation H is a typical inverse problem. Due to the conjugate symmetry of in-line digital holographic imaging, both the object and the twin image are solutions to this inverse problem. The main difference between them lies in sparsity, and CS is an effective method based on sparsity constraints. A common approach to determine ρ in Equation (8) is to solve the following minimization problem:

$$\min_{\rho} \frac{1}{2}\|I - G\rho\|_2^2 + \lambda R(\rho) \tag{9}$$

where G is the combined operator of the forward transformation F and the real part of the complex light field in Equation (8). Term $\|I - G\rho\|_2^2$ is the l2-norm of the residual used as a measure of reconstruction error. Term $R(\rho)$ is a regularization term which is typically used to constrain the solution space, imparting properties such as smoothness, sparsity, or other prior knowledge, and $\lambda > 0$ is a regularization parameter. In lensless digital holographic reconstruction, the total variation (TV) norm can be used as a measure of sparsity, which is defined by

$$\|\rho\|_{TV} = \sum_i \sqrt{\left|\Delta_i^x \rho\right|^2 + \left|\Delta_i^y \rho\right|^2} \tag{10}$$

where Δ_i^x and Δ_i^y represent the first-order local differential operators in the horizontal and vertical directions, respectively.

2.3. Architecture of ISTAHolo-Net

To solve the inverse problem in Equation (9), the proximal gradient descent method [28,29] is often used, and the iterative update step can be expressed by

$$\rho^k = \underset{\rho}{\mathrm{argmin}} \frac{1}{2}\left\|\rho - \left(\rho^{k-1} - \alpha \nabla g\left(\rho^{k-1}\right)\right)\right\|_2^2 + \lambda R(\rho) \tag{11}$$

where ρ^k represents the output after the k-th iteration, $g(\cdot) = \frac{1}{2}\|I - G\rho\|_2^2$ represents the data fidelity term, ∇ is the differential operator, and α is the step size. As a popular first-order proximal method, the iterative shrinkage thresholding algorithm solves the reconstruction problem in Equation (11) by iterating between the following update steps:

$$v^k = \rho^{k-1} - \alpha \nabla g\left(\rho^{k-1}\right) = \rho^{k-1} - \alpha G^{\mathrm{T}}\left(G\rho^{k-1} - I\right) \tag{12}$$

$$\rho^k = prox_{\lambda,R}(v^k) \tag{13}$$

where G^{T} is the adjoint operator of G. By iteratively updating v^k and ρ^k until convergence, the object image can be obtained. Solving $prox_{\lambda,R}(v^k)$ in an efficient and effective way

is critical for ISTA. Usually, ISTA requires many iterations to obtain a satisfactory result, suffering from extensive computation [29].

Inspired by the ISTA-Net, we propose a holographic image reconstruction network, dubbed the ISTAHolo-Net, whose architecture is illustrated in Figure 2, where the update steps are replaced by the cascaded sub-networks. Specifically, the two update steps shown in Equations (12) and (13) are achieved by two networks, which are named as the GDM and PMM, respectively. The ISTAHolo-Net consists of K phases, each of which corresponds to one iteration in the traditional ISTA.

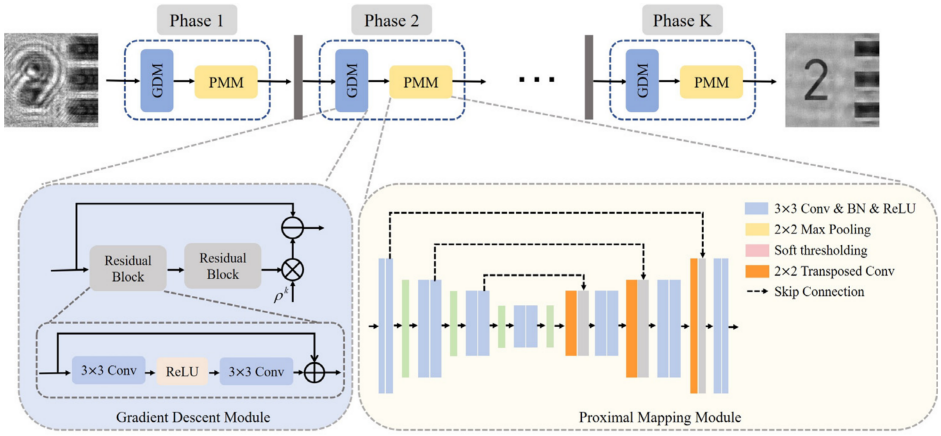


Figure 2. Schematic diagram of ISTAHolo-Net structure.

As shown in the blue box in Figure 2, the GDM consisting of two independent residual blocks is designed to obtain the residual term v^k where the gradient of the non-linear measurement expressed by Equation (8) should be known. Instead of calculating the gradient, we represent the imaging processes G^{T} and G in Equation (12) as $RB^k_{G^{\mathrm{T}}}(\cdot)$ and $RB^k_{G}(\cdot)$ with optimizable parameters. In addition, the step size α is set to be a learnable parameter. Therefore, the residual term can be expressed as:

$$v^k = \rho^{k-1} - \alpha^k RB^k_{G^{\mathrm{T}}}\left(RB^k_{G}\left(\rho^{k-1}\right) - I\right) \tag{14}$$

It can be seen that the primary function of $\rho^k = prox_{\lambda,R}(v^k)$ expressed by Equation (13) is to remove noise and artifacts from the intermediate results through a thresholding function. To effectively extract multi-scale features from images and consider the excellent performance of the U-Net on denoising tasks, we select the U-Net as the backbone network, where the final convolutional block is modified to a soft thresholding function (denoted by the green box in the PMM), forming the proximal mapping module. As illustrated in the yellow box in Figure 2, the PMM, which is based on the U-Net structure, consists of a down-sampling path and an up-sampling path. The down-sampling path of the U-Net is referred to as the encoder and is primarily used for feature extraction, denoted as $U(\cdot)$. The up-sampling path of the U-Net is referred to as the decoder and is primarily used for restoring image size, denoted as $\widetilde{U}$. Therefore, the proximal mapping module can be expressed by

$$\rho^k = \widetilde{U}\left(soft\left(U\left(v^k\right)\right), \theta\right) \tag{15}$$

where $soft(\cdot)$ represents the soft thresholding function, and θ represents its parameter. Assuming the input feature map is x_{in}, the soft thresholding function can be represented as:

$$soft(x_{in}, \theta) = sign(x_{in})Relu(|x_{in}| - \theta) \tag{16}$$

In the down-sampling path, we use a convolutional block with stride 3 and max-pooling with stride 2 at different levels, and the numbers of channels of the convolutional block from top to bottom are 16, 32, 64, and 128, respectively. In order to enlarge the image size, the up-sampling path includes a transposed convolutional layer between different layers. Except for the last layer, the up-sampling path and the down-sampling path are connected by a skip connection. In the last layer of the down-sampling path, the feature map passes a convolutional block with 256 channels; then, the output feature is processed by the soft threshold function before being input to the up-sampling path.

During the training process, the network parameters are first initialized, and the holographic image reconstructed by using ASM is input to the GDM at phase 1. After passing through the gradient descent module, feature map v^1 is obtained, which is then directly input into the PMM at phase 2 to produce the input for the next sub-network x^2. This process is repeated until all sub-networks have been traversed, at which point loss calculation and parameter updates are performed.

2.4. Loss Function

In the proposed ISTAHolo-Net, each sub-network can be regarded as one iteration of the traditional ISTA. To constrain the update process of each sub-network while simultaneously updating the overall network parameters, the objective function of this network is defined as a weighted sum of the mean squared errors between the reconstructed images of each sub-network and the corresponding label images.

$$\text{Loss} = \frac{\mu}{N_P - 1} \sum_{k=1}^{N_P - 1} \left\| y - D^k\left(x^k\right) \right\|_2^2 + \left\| y - D^{N_P}\left(x^{N_P}\right) \right\|_2^2 \tag{17}$$

where x^k represents the input image for each sub-network, y denotes the label image corresponding to the input holographic image, and $D^k\left(x^k\right)$ refers to the recovered image obtained after denoising by the k-th sub-network. The parameter μ is the weight assigned to the loss of the sub-network, set to 0.5 in this case to balance the reconstruction performance between the intermediate sub-networks and the final output sub-network. N_P represents the total number of sub-networks, which is set to 4 in this instance.

3. Results

3.1. System Introduction

To obtain experimental data, the LDH system is built, as shown in Figure 3. This system mainly consists of a laser, a pinhole, a CMOS image sensor, and a three-dimensional translation stage. The laser, pinhole, and image sensor are vertically aligned at the center, and all three are placed horizontally. The laser used is model CPS532-C2 from THORLABS company (Newton, NJ, USA), with an output power of 1.5 mW and a wavelength of 532 nm. The pinhole used has a diameter of 40 μm. The minimum step of the translation stage is 10 μm along the z direction. The MDX6-T CMOS image sensor (Mshot, Guangzhou, China), with a pixel size of 1.34 um and an effective pixel count of 16 million, provides an imaging resolution of 4608 × 3456 and supports an imaging field of view of up to 25.52 mm^2. To obtain a uniform plane wave illumination on the object, the initial distance between the laser and the sensor plane is controlled to be around 28 cm, and the initial distance from the sample to the sensor is within 1 mm. By adjusting the three-dimensional translation stage along z-axis, multiple holograms with different heights are collected for the same sample, with an interval of 10 μm between adjacent images. To avoid the noise from external light interference, the collection of holograms was conducted in a lightproof and enclosed environment.

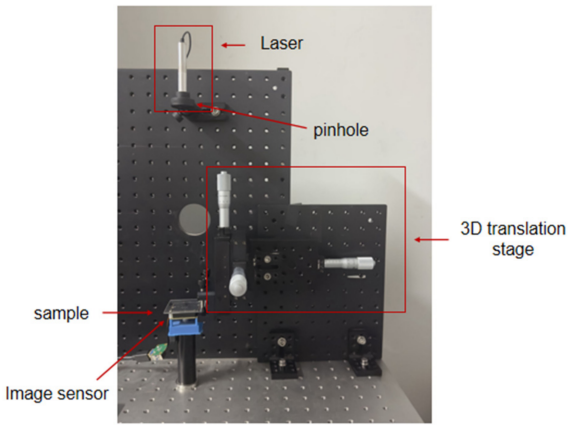


Figure 3. Hologram acquisition system.

3.2. Network Configuration and Parameter Settings

The training of the network is conducted on a computer with a 12th-Gen Intel (R) Core (TM) i5-12400F 2.50 GHz CPU, an NVIDIA GeForce RTX 3060 GPU with 12 GB of VRAM, and a Windows 64-bit operating system. The deep learning environments used are PyTorch 1.9.1, CUDA 11.1, and Python 3.7. During the training process, the Adam gradient descent algorithm is employed, with a total training count set to 200 iterations, an initial learning rate of 0.005, a step size initialized to 0.5 in the gradient descent module of each sub-network, and the soft threshold function parameter initialized to 0.01 in the proximal mapping module.

3.3. Datasets Preparation

To validate the performance of the proposed method, we constructed four simulation datasets to train and test the network. The simulation datasets consist of the handwritten MNIST dataset and a custom-made text dataset, as shown in Figure 4. The MNIST dataset, originating from the national institute of standards and technology of the United States, consists of the handwritten digits 0–9. Each image is black background with white digits, sized at 28 × 28 pixels. To match the network input requirements, each image was interpolated to 128 × 128 pixels. Sixty images were selected from each digit category. In total, 600 images served as the ground truth data. The custom-made text dataset is derived from the GB2312 Chinese character font set, including 6763 characters. Six hundred unique characters were selected from this set, and images with black backgrounds and white characters based on the selected characters were created. The size of the images is 128 × 128, and they are utilized as labels for the text dataset. Both datasets represent amplitude objects. Since lensless holographic imaging techniques are also frequently used for phase object imaging, we calculated all the label data from the above two kinds of datasets to generate the labels of the phase simulation datasets. The holograms of the dataset were simulated based on the optical diffraction theory expressed by Equations (5) and (6). The ratio of the data used for the training and test is 5:1 in our simulation experiment.

To validate the reconstruction capability of the network, the holograms of a USAF 1951 resolution chart and the EC109 cells (human esophageal cancer cells) were recorded. We used a USAF 1951 resolution chart and the EC109 cells (human esophageal cancer cells) as test samples in the experimental setup, as illustrated in Figure 5. For cell imaging, the cell suspension was first dropped onto a glass slide using a pipette. The cells were spread evenly, and the glass slide was placed directly on the image sensor. Holograms were captured using a computer-controlled camera. The ground truth images of the cells were obtained by using the multi-height holograms. To capture multi-height holograms, the sample was placed on a three-dimensional translation stage, with the initial position

adjusted carefully so that the sample was close to the image sensor surface. Then, to satisfy the requirements of the multi-height iterative reconstruction method, the translation stage was moved vertically with equal intervals, and a hologram was recorded at each interval. To determine the distance between the sample and the sensor, the auto-focusing function ToG [30] was calculated. For the cell image dataset construction, the compressive sensing-based reconstruction method was adopted to generate the label data. For the resolution chart dataset construction, the label data were made by using the multi-height iterative reconstruction method based on the hologram recorder with holographic images at six different heights. The original images captured were 4608 × 3456 pixels in size. To facilitate subsequent network training, the captured holograms were cropped down to 128 × 128 pixels. Poor-quality images were manually filtered out, and the remaining images formed the input holograms. Finally, the data were flipped, translated, and cropped, resulting in 3000 sets of images for each dataset. In the real experiment, to achieve better reconstruction results, the captured holograms were first back-propagated to the sample plane before being input into the network. The ratio of the training dataset to the test dataset is 2:1.

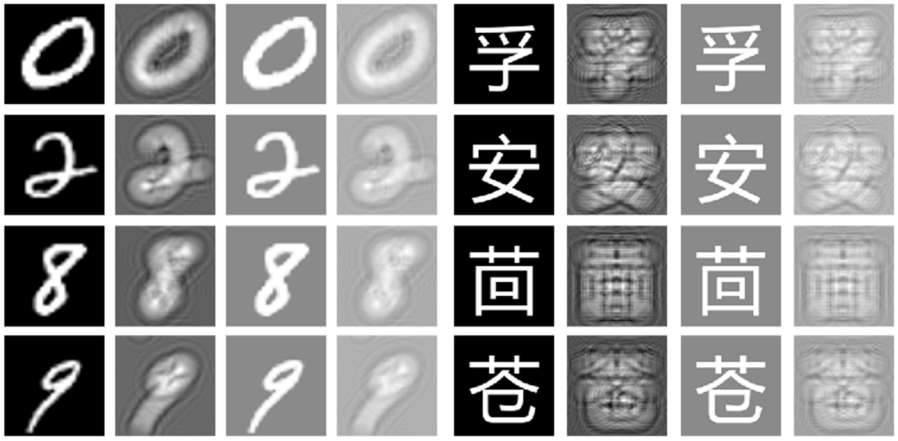


Figure 4. Datasets for simulation experiment: (**a**) amplitude MNIST data; (**b**) phase MNIST data; (**c**) amplitude text data; (**d**) phase text data. The left column corresponds to the ground truth, and the right column corresponds to the holograms.

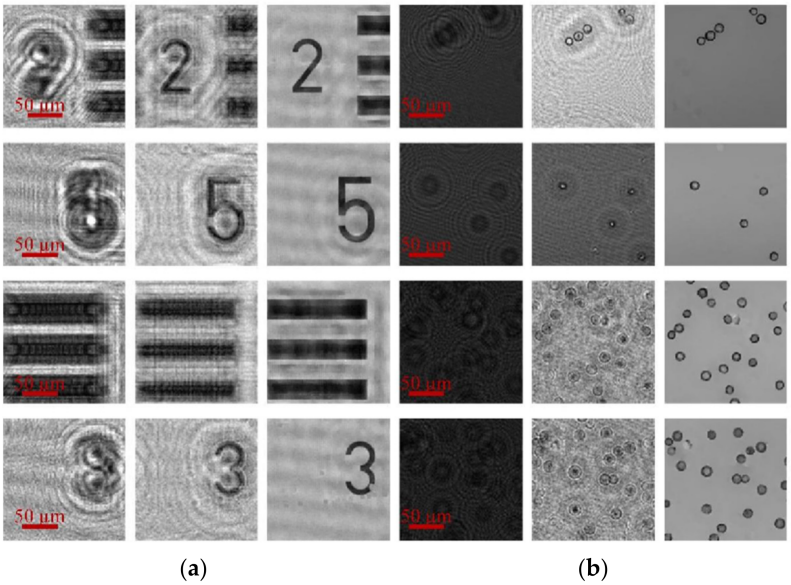


Figure 5. Datasets used in real experiment: (**a**) USAF resolution chart and (**b**) EC109 cell datasets. From left to right are recorded holograms, ASM results, and ground truth, respectively.

3.4. Evaluating Metrics

To quantify the reconstruction performance of the network, we use the peak signal-to-noise ratio (PSNR) [31] and structural similarity index measure (SSIM) [32] between the reconstructed images and the ground truth labels as evaluation metrics. The scores are obtained by calculating the average values over the reconstruction results of the test dataset.

The PSNR is defined based on mean square error (MSE) [33], which measures the average squared difference between corresponding pixels in the true and the reconstructed images. A higher MSE indicates greater dissimilarity between the images, reflecting poorer reconstruction quality of the network. MSE can be calculated by

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}|y(i,j) - x(i,j)|^2 \tag{18}$$

where m and n are the height and width of the two images, respectively, $x(i,j)$ and $y(i,j)$ represent the reconstructed result and the corresponding labeled image. The coordinates (i,j) denote the coordinates of the pixels in both images. The PSNR is defined as the logarithm of the ratio of the mean squared error between the original and processed images to $(2^n-1)^2$, expressed in decibels (dB). For grayscale images, n typically equals 8, which means the intensity is digitalized to 8 bits per pixel. The calculation formula for PSNR is defined as follows:

$$\mathrm{PSNR} = 10 \times \lg\left[\frac{(2^n-1)^2}{\mathrm{MSE}}\right] \tag{19}$$

It can be seen from the above equation that a larger PSNR indicates less distortion between the reconstructed image and the reference image, suggesting better reconstruction performance of the network. The PSNR is a commonly used metric for visualizing errors due to its simplicity and ease of computation. However, because human vision is not equally sensitive to all types of errors, PSNR scores may not always correlate perfectly with human perception. Therefore, this paper introduces another widely used image quality assessment metric, the SSIM, which differs from the PSNR by not only measuring absolute errors between pixel values but also integrating the brightness, contrast, and structural features of the images. The SSIM provides a more intuitive evaluation of the degree of distortion and similarity in reconstructed results, aligning better with human visual perception. The specific calculation formula for the SSIM is as follows:

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{20}$$

where x and y represent the output result of the network and its corresponding reference image. $l(x,y)$ is the luminance comparison function, $c(x,y)$ is the contrast comparison function, and $s(x,y)$ is the structure comparison function. α, β, and γ are weights corresponding to the luminance, contrast, and structure functions, respectively. These functions are specifically defined as follows:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{21}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{22}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{23}$$

where μ_x and μ_y represent the mean values of images x and y, respectively, σ_x and σ_y represent the standard deviations of images x and y, respectively, and σ_{xy} represents the

covariance between images x and y. For practical applications, α, β, and γ are typically set to 1. Additionally, to avoid instability in the results due to the denominator being too close to zero, three constants, C_1, C_2, and C_3, are introduced. $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, $C_3 = C_2/2$. Typically, $K_1 = 0.01$, $K_2 = 0.03$, and L represents the maximum value of the pixels. Further derivation yields the following:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\mu_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{24}$$

It can be observed from the above equations that the SSIM metric ranges from 0 to 1. A value closer to 1 indicates greater similarity between the two images, whereas a lower value suggests increasing dissimilarity between them. This paper assesses the reconstruction performance of the model using both the PSNR and SSIM metrics, aiming to provide an objective evaluation that aligns with subjective human perception, thus ensuring comprehensive and reliable assessment.

3.5. Experimental Results and Analysis

3.5.1. Reconstruction Results of Simulated Amplitude Object

In this section, we evaluate the performance of our proposed method on the simulated amplitude object and phase object dataset. We also compare the reconstruction results with those obtained by the ASM, HRNet, Y-Net, DH-GAN, and ISTAHolo-Net. Figure 6 illustrates the reconstruction results for the amplitude MNIST dataset and the amplitude text dataset. All the results are normalized to the largest pixel value in the ground truth image. It can be observed that the ASM can roughly reconstruct the boundary contours of the target symbols, but it performs poorly in terms of the overall color reconstruction, with dense twin images surrounding the original images and in the background. The reconstruction results of the DL networks are better than those of the ASM. Notably, the HRNet reconstruction results exhibit slight twin-image interference. The Y-Net obtains the sharpest results while the digitals are a little distorted. The DH-GAN is able to recover the digits with high background especially when is target is dense. Few twin images show up in the reconstruction results obtained by the ISTAHolo-Net for the amplitude MNIST object. However, the output of the ISTAHolo-Net is less sharp than that of the Y-Net, which may be caused by the fact that the trace of the digitals is rather smooth while the ISTAHolo-Net uses the TV regularization.

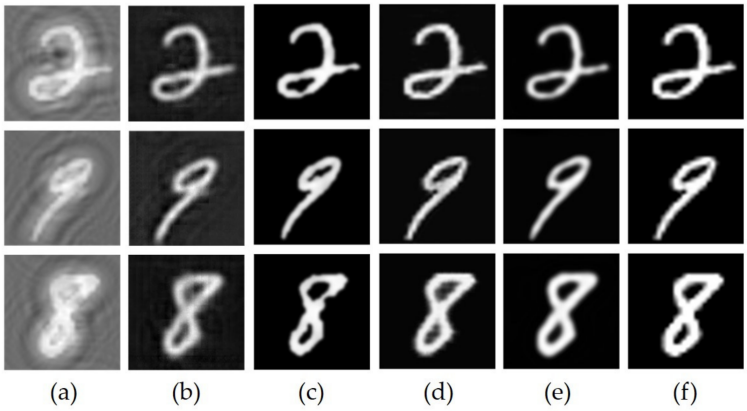


Figure 6. Reconstruction results for amplitude MNIST dataset: (**a**) ASM, (**b**) HRNet, (**c**) Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

For the simulated text dataset, the reconstructed images are shown in Figure 7. All the results are normalized to the largest pixel value in the ground truth image. The results

obtained by the ASM are blurred seriously by the twin image. The HRNet can recover the character with little twin-image interference and thus has high background. The Y-Net performs better than the HRNet, while it is a little worse than the DH-GAN and ISTAHolo-Net. The DH-GAN obtains clear characters with high background. The ISTAHolo-Net obtains the clearest and sharpest results, which are closest to the labeled images among the four methods, demonstrating the effectiveness of the proposed network.

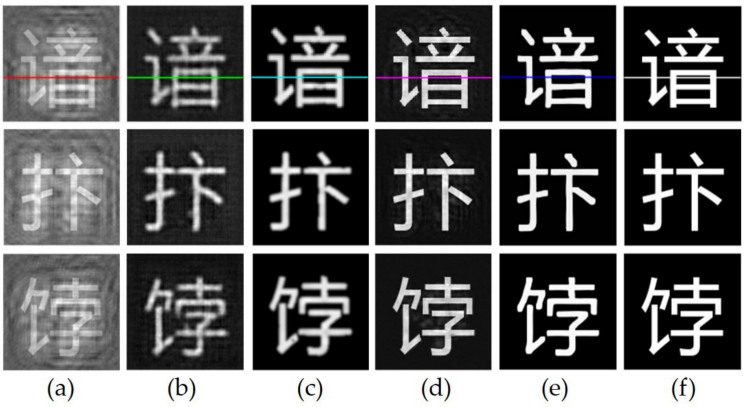


Figure 7. Reconstruction results for amplitude text dataset: (**a**) ASM, (**b**) HRNet, (**c**) Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

To better visually demonstrate the effectiveness of the ISTAHolo-Net, the normalized profiles across the red, green, cyan, purple, blue, and gray lines for a random selected sample in the text dataset are illustrated in Figure 8, from which we can see that the profile of the ISTAHolo-Net reconstruction matches the ground truth best.

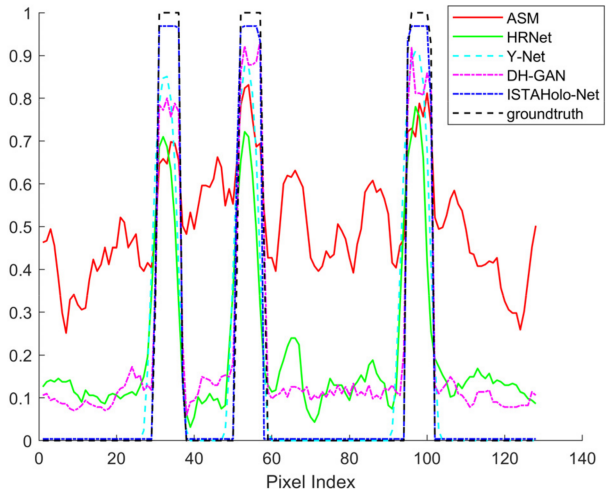


Figure 8. Normalized gray level profiles along the lines in Figure 7.

To quantitatively evaluate the performance of the reconstruction methods, the PSNR and SSIM are calculated and are listed in Table 2. It can be seen that the SSIM and PSNR scores of the ASM are the lowest. The HRNet recovers the structural information of the images relatively well, but the reconstruction results have a significant contrast difference compared to the labeled images, resulting in a lower SSIM score for the HRNet, although its PSNR score is higher. The Y-Net gets a better result than the ASM and HRNet. The ISTAHolo-Net achieves the highest SSIM and PSNR scores among these methods, and comparisons of the ISTAHolo-Net's scores across the two datasets indicate that it maintains a consistently excellent reconstruction performance when reconstructing different types of samples.

Table 2. Quantitative metrics for the amplitude target reconstruction using ASM, HRNet, Y-Net, DH-GAN, and ISTAHolo-Net.

Amplitude Object	Metrics	ASM	HRNet	Y-Net	DH-GAN	ISTAHolo-Net
MNIST dataset	PSNR (dB)	7.38	19.38	20.74	22.45	23.18
	SSIM	0.31	0.32	0.89	0.53	0.91
Text dataset	PSNR (dB)	7.57	13.85	16.08	20.06	24.49
	SSIM	0.27	0.35	0.81	0.57	0.90

3.5.2. Reconstruction Results of Simulated Phase Objects

The recovered images of the simulated phase objects by different methods are illustrated in Figures 9 and 10. The reconstruction performance of the five algorithms is consistent with their performance in amplitude object experiments. Due to the presence of residual structures and limited training data, the HRNet network converges slowly, and the learned information is incomplete, resulting in black gaps in the reconstructed images. Still, the ASM is affected by the twin-image interference. The performance of the Y-Net is close to that of the ISTAHolo-Net. The ratio of the signal to the background is highest for the DH-GAN approach. Compared with the other algorithms, the ISTAHolo-Net introduces the least noise in the reconstructed image and is the closest to the ground truth, which is also demonstrated by the normalized gray level profiles shown in Figure 11.

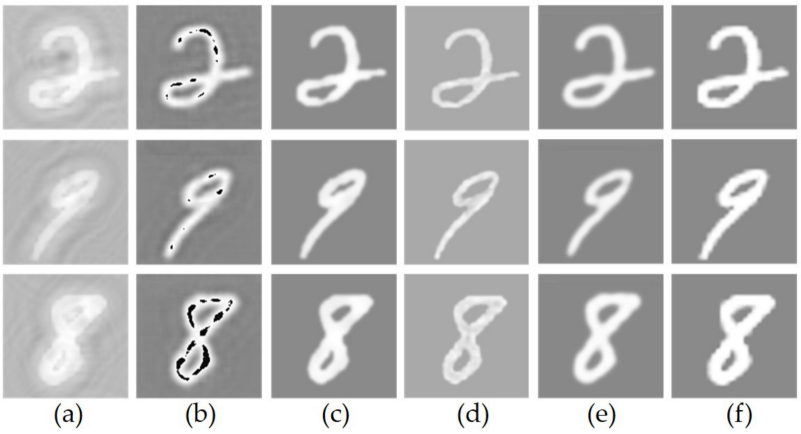


Figure 9. Reconstruction results of phase target by different methods: (**a**) ASM, (**b**) HRNet, (**c**) Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

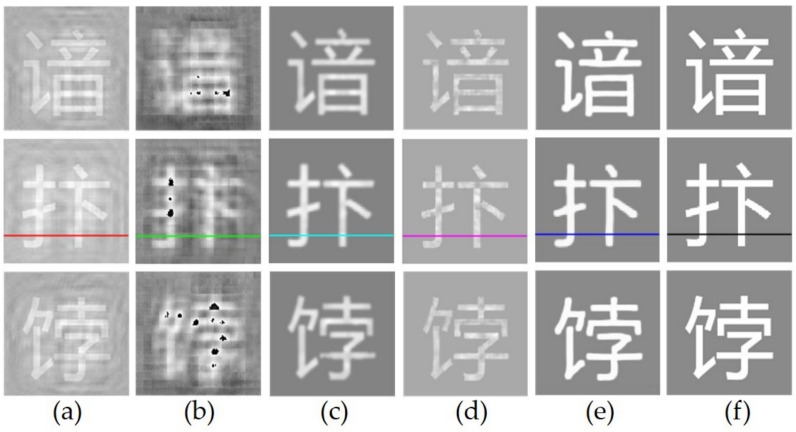


Figure 10. Recovered phase text dataset by different methods: (**a**) ASM, (**b**)HRNet, (**c**)Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

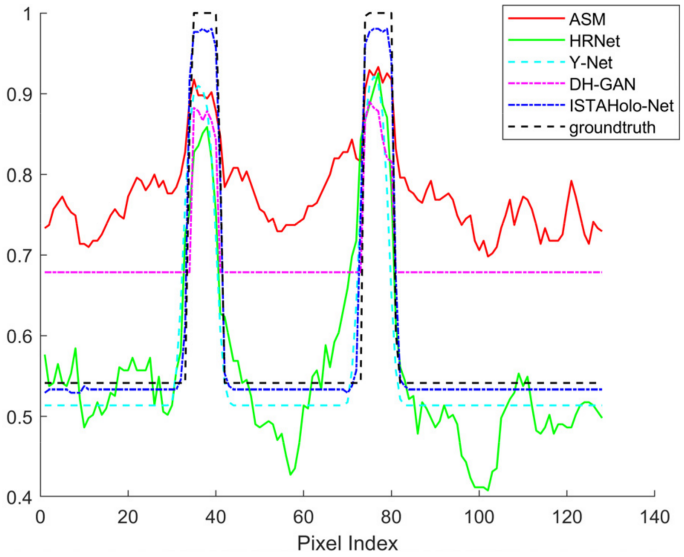


Figure 11. Normalized profiles along the red, green, cyan, purple, blue, and black lines in Figure 10.

Table 3 presents the numerical evaluation metrics of the recovered results of the phase objects. As shown in the table, the performances of the ASM, Y-Net, and HRNet for the phase images are better compared to those for the amplitude objects. Outperforming the other methods, the ISTAHolo-Net achieves SSIM scores above 0.9 and PSNR scores above 23 dB for both the amplitude and phase targets.

Table 3. Quantitative metric values for the phase target reconstruction using ASM, HRNet, and ISTAHolo-Net.

Phase Object	Metrics	ASM	HRNet	Y-Net	DH-GAN	ISTAHolo-Net
MNIST dataset	PSNR (dB)	14.54	20.77	23.07	18.57	23.27
	SSIM	0.64	0.67	0.92	0.90	0.94
Text dataset	PSNR (dB)	15.03	17.26	21.16	16.80	24.55
	SSIM	0.67	0.51	0.85	0.82	0.92

3.5.3. Reconstruction Results of Real Objects

To further investigate the performance of the network, real experiments were conducted, where the holograms of the resolution chart and the appended EC109 cells were recorded. The recovered results are shown in Figures 12 and 13. The reconstruction results of the ASM exhibit significant twin images, particularly for cell samples, which severely affects the observation of image information. The HRNet achieves relatively clear reconstruction results when reconstructing the resolution chart; however, its performance significantly declines when reconstructing cells. Still, the Y-Net performs better than the ASM and the HRNet. The ISTAHolo-Net provides the best reconstruction results for both the resolution chart and the EC109 cells. It effectively smooths the background of images and highlights detailed information within the images, which is beneficial for the subsequent observation and application of the reconstruction results. Furthermore, as illustrated in Figure 14, in the resolution chart experiment, the normalized profile of the ISTAHolo-Net reconstruction is closer and smoother compared to those of other reconstruction methods, which means that the reconstructed result images contain the least amount of noise and twin-image interference.

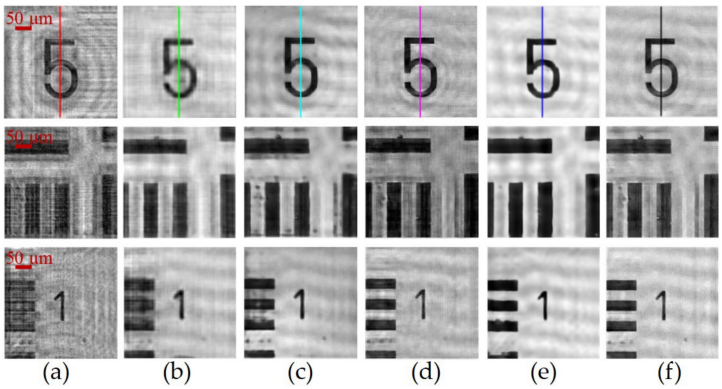


Figure 12. Reconstruction results for USAF resolution chart: (**a**) ASM, (**b**) HRNet, (**c**) Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

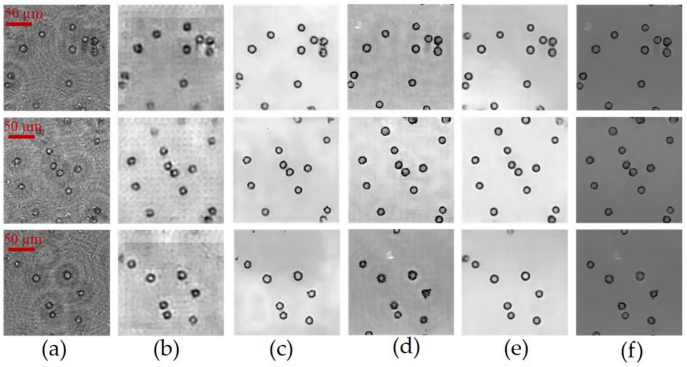


Figure 13. Reconstruction results for EC109 cells by different methods: (**a**) ASM, (**b**) HRNet, (**c**) Y-Net, (**d**) DH-GAN, (**e**) ISTAHolo-Net, (**f**) ground truth.

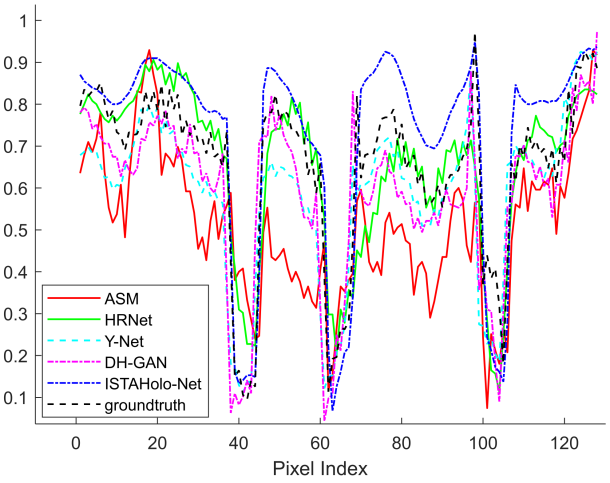


Figure 14. Profiles of the recovered USAF resolution chart by different methods.

Table 4 presents the metric values obtained by different reconstruction methods on the real data. As shown in the table, for the USAF resolution chart, which is a typical amplitude object, the ISTAHolo-Net exhibits the highest PSNR and SSIM scores. While for the appended cells, which are typical phase objects, the PSNR value obtained by the proposed approach ranks at fourth place, although the SSIM is higher than those obtained by other methods. It is surprising that the ASM achieves the highest PSNR value. That is caused by the fact that only several cells are sparsely located in the liquid culture medium, and hence, most of the pixels in the image are occupied by the background. As the background values are similar in Figure 13, the ASM obtains the largest PSNR, which measures the averaged gray value difference of all the pixels in the two images. The results

also suggest that the contrast of the cell against the background is enhanced by all the DL-based algorithms.

Table 4. Comparison of reconstruction performance for the real datasets among ASM, HRNet, Y-Net, and ISTAHolo-Net.

Object	Metrics	ASM	HRNet	Y-Net	DH-GAN	ISTAHolo-Net
USAF	PSNR (dB)	15.39	20.29	20.55	20.84	22.43
	SSIM	0.51	0.69	0.81	0.79	0.81
Cell	PSNR (dB)	16.29	9.42	6.90	11.69	7.22
	SSIM	0.40	0.62	0.75	0.77	0.89

3.6. Ablation Study

As mentioned in Section 2.3, there are K phases in the proposed network. In this section, we conduct an ablation experiment to determine the optimal value of K. The number of sub-networks composing the ISTAHolo-Net is set to different values which are 2, 3, 4, and 5, respectively, and the reconstructed results for the resolution target dataset are shown in Figure 15. The results indicate that when K = 4, the restoration of image details is the best, closely matching the ground truth. Additionally, we calculate the SSIM and PSNR of the resolution target reconstruction results for different values of K, as shown in Figure 16. The PSNR values for K = 2, K = 3, K = 4, and K = 5 are 16.18, 18.42, 22.43, and 17.14 dB, respectively. The SSIM values for K = 2, K = 3, K = 4, and K = 5 are 0.57, 0.62, 0.81, and 0.60, respectively. It can be seen that increasing K improves image quality when K < 5. However, the image quality decreases if the number of stages in the ISTAHolo-Net continues to increase. Therefore, the number of sub-networks K was set to 4 in all the experiments.

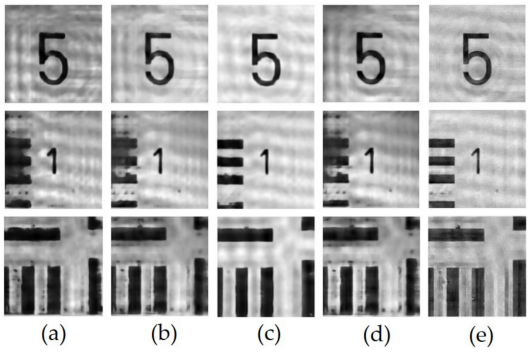


Figure 15. USAF resolution target reconstruction: (**a**) K = 2; (**b**) K = 3; (**c**) K = 4; (**d**) K = 5; (**e**) ground truth.

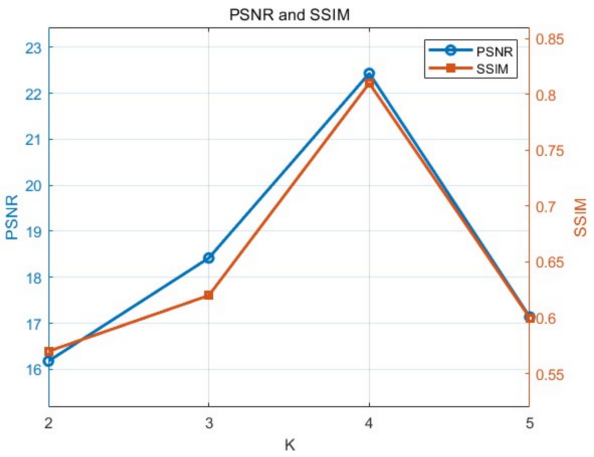


Figure 16. PSNR and SSIM for resolution chart reconstruction with different K values (2, 3, 4, and 5).

4. Discussion

In LDH reconstruction, the object wave field should be recovered from the measured hologram. The ASM has been used widely for LDH reconstruction as it is analytical and can be implemented quickly through fast Fourier transform. However, the twin image shows up in the recovered object image as only the intensity can be recorded by the CMOS or CCD sensor. To remove the twin image, the ISTA can be used to solve the inverse problem in LDH reconstruction. As an iterative algorithm, the ISTA solves the regularized reconstruction problem with sparsity as a prior. It requires a large number of iterations to approach the optimal solution, and the selection of parameters and the optimal transform needs to be predefined.

In this paper, the ISTAHolo-Net is proposed for LDH reconstruction, in which the unfolding of a deep-learning-based network maps the update steps in a traditional ISTA-based holographic reconstruction to a neural network architecture with a fixed number of stages. In addition, the step size and regularization weight are set to be learnable parameters in the proposed approach. To determine the number of stages in the proposed network, the ablation study is conducted. The results show that the image quality is improved with increasing K, but it decreases if K continues to increase, and the best image quality is achieved when K = 4. This may be caused by the fact that the number of unknown parameters increases with larger K.

To evaluate the effectiveness of the proposed method, the simulation experiments for both amplitude and phase objects were conducted. The PSNR and SSIM were calculated. From the image reconstruction results and evaluation metric scores, the reconstruction performance of the ISTAHolo-Net outperforms the traditional ASM, HRNet, Y-Net, and DH-GAN. The reconstruction results of the ASM suffer from significant twin-image interference, leading to poor overall image quality and the lowest score. The HRNet and Y-Net are able to reduce the twin-image interference to some extent, but some image information is lost during the reconstruction process, resulting in blurred images with poor contrast. In the simulation experiment, the background in the recovered image obtained by the DH-GAN is highest. Additionally, the HRNet has higher data requirements for training, and under the same experimental conditions, the black artifacts it produces severely affect the reconstruction results, although it performs better than the ASM. In terms of metric scores, the HRNet shows a significant improvement in the PSNR over the ASM, but the SSIM score improves only slightly and even drops below that of the ASM. On the other hand, the ISTAHolo-Net achieves the best reconstruction results and the highest scores for both amplitude and phase objects in the simulation experiments, and this excellent performance can be consistently maintained across different datasets, demonstrating high generalization capability.

In addition, the parameters of the networks used in the simulation experiment are listed in Table 5. From the table, we can see that the ISTAHolo-Net has about 1.9 M parameters and that the test time is about 10 ms, which is the fastest among the four approaches.

Table 5. Parameters of the networks.

	HRNet	**Y-Net**	**DH-GAN**	**ISTAHolo-Net**
Parameters (M)	2.85	16	0.9	1.9
Test Time (ms)	20	75	85	10
Model Size (MB)	28	30	67	26

Furthermore, real experiments were conducted to explore the ISTAHolo-Net. The holograms of the USAF resolution chart and the EC109 cell lines were recorded using the LDH system. As observed, due to interference from some uncertain factors in the real

experiments and the increased complexity of the samples, the reconstruction results of all the methods were weakened compared with the simulation experiments. However, our proposed network still produced the best reconstruction results among the five methods. It is worth mentioning that the PSNR values of the cell reconstruction results of the DL-based approaches are lower than those of the ASM, which is caused by the fact that only several cells are observed, and most of the pixels in the images are occupied by the background, which contributes to the final PSNR metric value. Fortunately, the contrast of the cell against the background is enhanced by all the DL-based reconstruction methods, leading to improved image quality.

While the proposed method shows promising results in lensless digital holographic reconstruction, there are several limitations that should be addressed in future work. Firstly, the ISTAHolo-Net is a fully supervised reconstruction network, and it is highly dependent on the quality of the training dataset. Low-quality label data may lead the network to learn incorrect patterns, and in situations where a large amount of labeled data is required or where labeling is costly, the network's performance may degrade. Therefore, evolving the ISTAHolo-Net into an unsupervised reconstruction network is a promising direction to address cases where the labels are difficult to obtain. In addition, the ISTAHolo-Net decomposes the traditional ISTA model into deep learning network modules; its reconstruction performance is somewhat restricted to the algorithm itself, and its interpretability still needs to be explored. As a result, future work could also focus on improving image reconstruction quality by optimizing the algorithm. Thirdly, the performance of the proposed method was only tested by holographic reconstruction of simulation data, the resolution chart, and suspended cells. A more diverse set of test images, mainly from real-world conditions, should be tested. We will apply the method to stainless pathological section and live cell imaging in the future. Finally, the training of the network is time-consuming. The training time is about 4.1 h for 500 128×128 holograms. With the trained parameters, the reconstruction time is about 8 ms for one hologram while the reconstruction time is about 1 ms when using the ASM.

5. Conclusions

In this paper, we propose a deep unfolding network for LDH reconstruction with higher interpretability; it is dubbed the ISTAHolo-Net and is based on the inverse model with sparsity as a prior. To resolve the inverse problem, the ISTAHolo-Net replaces the iterative update steps in ISTA method with a fixed number of phases or sub-networks. Each phase consists of two modules, the GDM and PMM, respectively. The GDM is used to generate the immediate reconstruction result, while the PMM aims to finish the proximal mapping based on the immediate reconstruction result. By modularizing the mathematical formulas into neural network components, we avoid interference from physical disturbances in the experimental results. Through several training iterations, the network parameters are continuously updated to improve the holographic image reconstruction performance. To validate the network's feasibility, we constructed datasets with different samples and compared the recovered object images using the proposed method with the ASM, HRNet, Y-Net, and DH-GAN approaches. In both the amplitude and phase object simulation experiments, the ISTAHolo-Net outperforms the other methods in terms of reducing twin-image artifacts and recovering details. The PSNR scores are all above 23 dB, and the SSIM scores are no less than 0.9, far exceeding the performance of the other methods. We then conducted real experiments, where the holograms of the USAF resolution target and EC109 cells were recorded and the samples were recovered. The results show that in more challenging reconstruction tasks, the ASM's reconstructions exhibit severe twin-image artifacts that hinder image information interpretation. The

HRNet, Y-Net, and DH-GAN can recover the object wave field better than the ASM. The ISTAHolo-Net achieves the highest SSIM score among the five methods compared. However, due to the large difference between the background of the reconstructed image and the labels, the ISTAHolo-Net obtains the second lowest PSNR score. In the future, we will test our method with a more diverse set of samples, such as stainless pathological sections and live cells.

Author Contributions: Conceptualization, D.C. and X.C.; writing—original draft preparation, D.C. and H.G.; writing—review and editing, Z.G. and D.C.; funding acquisition, D.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Research and Development Program of Shaanxi (Program No. 2023-YBSF-258).

Data Availability Statement: The data are available upon reasonable request to interested researchers.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Seo, S.; Su, T.-W.; Tseng, D.K.; Erlinger, A.; Ozcan, A. Lensfree holographic imaging for on-chip cytometry and diagnostics. *Lab A Chip* **2009**, *9*, 777–787. [CrossRef] [PubMed]
2. Wu, Y.; Ozcan, A. Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring. *Methods* **2018**, *136*, 4–16. [CrossRef] [PubMed]
3. Kreis, T.M.; Adams, M.; Jüptner, W.P. Methods of digital holography: A comparison. In *Optical Inspection and Micromeasurements II*; SPIE: San Francisco, CA, USA, 1997.
4. Chen, D.; Wang, L.; Luo, X.; Xie, H.; Chen, X. Resolution and Contrast Enhancement for Lensless Digital Holographic Microscopy and Its Application in Biomedicine. *Photonics* **2022**, *9*, 358. [CrossRef]
5. Guo, C.; Liu, X.; Kan, X.; Zhang, F.; Tan, J.; Liu, S.; Liu, Z. Lensfree on-chip microscopy based on dual-plane phase retrieval. *Opt. Express* **2019**, *27*, 35216–35229. [CrossRef]
6. Guo, Y.; Guo, R.; Qi, P.; Zhou, Y.; Zhang, Z.; Zheng, G.; Zhong, J. Robust multi-angle structured illumination lensless microscopy via illumination angle calibration. *Opt. Lett.* **2022**, *47*, 1847–1850. [CrossRef]
7. Zhang, W.; Cao, L.; Brady, D.J.; Zhang, H.; Cang, J.; Zhang, H.; Jin, G. Twin-image-free holography: A compressive sensing approach. *Phys. Rev. Lett.* **2018**, *121*, 093902. [CrossRef]
8. Kim, H.; Song, G.; You, J.I.; Lee, C.; Jang, M. Deep learning for lensless imaging. *J. Korean Phys. Soc.* **2022**, *81*, 570–579. [CrossRef]
9. Rivenson, Y.; Zhang, Y.; Günaydın, H.; Teng, D.; Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **2018**, *7*, 17141. [CrossRef]
10. Wang, H.; Lyu, M.; Situ, G. eHoloNet: A learning-based end-to-end approach for in-line digital holographic reconstruction. *Opt. Express* **2018**, *26*, 22603–22614. [CrossRef]
11. Ren, Z.; Xu, Z.; Lam, E.Y. End-to-end deep learning framework for digital holographic reconstruction. *Adv. Photonics* **2019**, *1*, 016004. [CrossRef]
12. Wang, K.; Dou, J.; Kemao, Q.; Di, J.; Zhao, J. Y-Net: A one-to-two deep learning framework for digital holographic reconstruction. *Opt. Lett.* **2019**, *44*, 4765–4768. [CrossRef] [PubMed]
13. Yin, D.; Gu, Z.; Zhang, Y.; Gu, F.; Nie, S.; Ma, J.; Yuan, C. Digital holographic reconstruction based on deep learning framework with unpaired data. *IEEE Photonics J.* **2020**, *12*, 3900312. [CrossRef]
14. Huang, L.; Liu, T.; Yang, X.; Luo, Y.; Rivenson, Y.; Ozcan, A. Holographic image reconstruction with phase recovery and autofocusing using recurrent neural networks. *ACS Photonics* **2021**, *8*, 1763–1774. [CrossRef]
15. Chen, H.; Huang, L.; Liu, T.; Ozcan, A. Fourier imager network (FIN): A deep neural network for hologram reconstruction with superior external generalization. *Light Sci. Appl.* **2022**, *11*, 254. [CrossRef]
16. Kiriy, S.A.; Rymov, D.A.; Svistunov, A.S.; Shifrina, A.V.; Starikov, R.S.; Cheremkhin, P.A. Generative adversarial neural network for 3D-hologram reconstruction. *Laser Phys. Lett.* **2024**, *21*, 045201. [CrossRef]
17. Wang, F.; Bian, Y.; Wang, H.; Lyu, M.; Pedrini, G.; Osten, W.; Barbastathis, G.; Situ, G. Phase imaging with an untrained neural network. *Light Sci. Appl.* **2020**, *9*, 77. [CrossRef]
18. Zhang, X.; Wang, F.; Situ, G. BlindNet: An untrained learning approach toward computational imaging with model uncertainty. *J. Phys. D Appl. Phys.* **2022**, *55*, 034001. [CrossRef]
19. Chen, X.; Wang, H.; Razi, A.; Kozicki, M.; Mann, C. DH-GAN: A physics-driven untrained generative adversarial network for holographic imaging. *Opt. Express* **2023**, *31*, 10114–10135. [CrossRef]

20. An, Q.; Liu, X.; Men, G.; Dou, J.; Di, J. Frequency-domain learning-driven lightweight phase recovery method for in-line holography. *Opt. Express* **2025**, *33*, 5890–5899. [CrossRef]
21. Zhang, Y.; Andreas Noack, M.; Vagovic, P.; Fezzaa, K.; Garcia-Moreno, F.; Ritschel, T.; Villanueva-Perez, P. PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets. *Opt. Express* **2021**, *29*, 19593–19604. [CrossRef]
22. Tian, Z.; Ming, Z.; Qi, A.; Li, F.; Yu, X.; Song, Y. Lensless computational imaging with a hybrid framework of holographic propagation and deep learning. *Opt. Lett.* **2022**, *47*, 4283–4286. [CrossRef] [PubMed]
23. Lee, C.; Song, G.; Kim, H.; Ye, J.C.; Jang, M. Deep learning based on parameterized physical forward model for adaptive holographic imaging with unpaired data. *Nat. Mach. Intell.* **2023**, *5*, 35. [CrossRef]
24. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; p. 04597.
25. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. *Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks*; IEEE: Piscataway, NJ, USA, 2017; pp. 2242–2251.
26. Zhang, Y.; Ritschel, T.; Villanueva-Perez, P. Reusability report: Unpaired deep-learning approaches for holographic image reconstruction. *Nat. Mach. Intell.* **2024**, *6*, 284–290. [CrossRef]
27. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018.
28. Wu, Y.; Cheng, H.; Wen, Y. Suppression of zero-order term and twin-image in in-line digital holography by a single hologram. *Meas. Sci. Technol.* **2020**, *31*, 025204. [CrossRef]
29. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [CrossRef]
30. Memmolo, P.; Distante, C.; Paturzo, M.; Finizio, A.; Ferraro, P.; Javidi, B. Automatic focusing in digital holography and its application to stretched holograms. *Opt. Lett.* **2011**, *36*, 1945–1947. [CrossRef]
31. Horé, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the IEEE International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
32. Wang, Z. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]
33. Sara, U.; Akter, M.; Uddin, M.S. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study. *J. Comput. Commun.* **2019**, *7*, 8–18. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.