

Learning from better simulation: creating highly realistic synthetic data for deep learning in scattering media

Bozhen Zhou^a, Zhitao Hao^a, Zhenbo Ren^{b,c,*}, Edmund Y. Lam^d, Jianshe Ma^a, and Ping Su^{a,*}

^aTsinghua University, Tsinghua Shenzhen International Graduate School, Shenzhen, China

^bNorthwestern Polytechnical University, School of Physical Science and Technology, Xi'an, China

^cNorthwestern Polytechnical University, Shenzhen Research Institute, Shenzhen, China

^dThe University of Hong Kong, Department of Electrical and Electronic Engineering, Hong Kong, China

Abstract. Obtaining the ground truth for imaging through the scattering objects is always a challenging task. Furthermore, the scattering process caused by complex media is too intricate to be accurately modeled by either traditional physical models or neural networks. To address this issue, we present a learning from better simulation (LBS) method. Utilizing the physical information from a single experimentally captured image through an optimization-based approach, the LBS method bypasses the multiple-scattering process and directly creates highly realistic synthetic data. The data can then be used to train downstream models. As a proof of concept, we train a simple U-Net solely on the synthetic data and demonstrate that it generalizes well to experimental data without requiring any manual labeling. 3D holographic particle field monitoring is chosen as the testing bed, and simulation and experimental results are presented to demonstrate the effectiveness and robustness of the proposed technique for imaging of complex scattering media. The proposed method lays the groundwork for reliable particle field imaging in high concentration. The concept of utilizing realistic synthetic data for training can be significantly beneficial in various deep learning-based imaging tasks, especially those involving complex scattering media.

Keywords: scattering medium; computational imaging; simulator training; digital holography; 3D particle field.

Received Mar. 27, 2025; revised manuscript received Jul. 2, 2025; accepted for publication Jul. 14, 2025; published online Aug. 26, 2025.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.APN.4.5.056007](https://doi.org/10.1117/1.APN.4.5.056007)]

1 Introduction

Optical imaging through scattering media,¹ which includes elements such as turbid water,² fog, and biological tissues,³ is vastly applicable across various fields. These range from high-speed color imaging⁴ and optical coherence tomography⁵ to quantitative phase imaging⁶ and biomedical imaging.^{7,8} However, the process of getting a deeper, clearer view of the details is predominantly hindered by scattering, particularly the multiple scattering caused by the complexity and inhomogeneity of tissues.⁹ Recently, deep learning techniques have emerged as critical tools for addressing these complex imaging

problems, and they have been successfully applied across various imaging disciplines.¹⁰ These include phase imaging,^{11,12} auto-focusing,¹³ and super-resolution imaging.^{14,15} There is also a lot of work devoted to the specific area of imaging through scattering media.^{16–18}

In the field of computational imaging,¹⁹ although deep learning has made significant strides in addressing imaging challenges, its dependency on large-scale training data presents a crucial issue: the acquisition of ground truth labels for model training is a cumbersome and time-consuming task. In certain domains, these labels can be obtained through established gold standards;²⁰ however, this approach is not only inefficient but also becomes impractical when the volume of data scales into the millions. Furthermore, in some fields, attaining ground truth remains an elusive goal due to the absence of effective solutions

*Address all correspondence to Zhenbo Ren, zbren@nwpu.edu.cn; Ping Su, su.ping@sz.tsinghua.edu.cn

to overcome specific technical challenges. For example, in particle field monitoring, acquiring ground truth for real, dynamic particle fields is always nearly impossible,²¹ especially at high particle concentrations where multiple scattering and defocusing significantly interfere with particle imaging. Many researchers^{22–25} modified the raw reconstructed images or scanning microscope images via digital image processing methods, as “ground truth.” This approach, however, lacks stability and accuracy and often requires significant time to complete the processing. Moreover, although “ground truth” derived from low concentrations may be somewhat reliable, its credibility diminishes at higher concentrations. Thus, efficiently obtaining sufficient data for deep learning training within a brief time-frame, while ensuring credible ground truth, stands as a pivotal challenge in utilizing deep learning for imaging through scattering media.

A promising strategy entails the use of synthetic data for network training while guaranteeing that the network retains the ability to process experimental data effectively. Domain adaptation from simulation to reality²⁶ is currently primarily accomplished through two methodologies: data manipulation and model manipulation.²⁷ Some researchers have successfully generated realistic data efficiently in specific domains, such as the generation of ultrasonic multimodal total focusing method (M-TFM) images.^{28,29} Particularly, in particle holography, however, due to the extreme complexity in the multiple-scattering process,^{30,31} a highly realistic synthetic data can be quite difficult to generate. In the field of particle holography, current research focuses on model manipulation. Chen et al.²¹ introduced a simulation-training model Holo-Net, which incorporates the physical prior of the point spread function (PSF), enhancing the network’s stability in particle localization and tracking tasks. However, the exclusive reliance on PSF prior introduces certain limitations when addressing more complex particle fields. Therefore, their experimental validation involved a

particle field of low concentration (~ 50 particles/mL) and relatively significant errors (~ 1 mm in the axial direction). Tahir et al.³² developed an extensive synthesized network, comprising multiple deep neural networks, to capture the diversity in the data and a substantial training set to mitigate the risk of overfitting, ensuring adaptability across various scattering conditions. Both of them need a substantial amount of data for training an extensive model, which is quite demanding on the expensive computational cost.

Diverging from their approaches, we manipulate the data and introduce a method termed “learning from better simulation” (LBS). This method greatly reduces the required amount of data for a modified U-Net model by enhancing the synthetic data used in training to closely mirror real data characteristics. As a result, the model can effectively handle experimental data with high particle concentrations. As shown in Fig. 1(a), an experimental hologram of the particle field is captured and subsequently backpropagated to various distances to acquire raw reconstruction images. By feeding these images into the simulator-trained U-Net as the input, we can get the required 2D layer images as the output. After applying the output processing algorithm to the output, a 3D particle field is reconstructed. Figure 1(b) shows how we get better synthetic training images for the U-Net by utilizing experimental images. At step 1, we use an optimization algorithm to continuously match the reference experimental image from an experimental hologram with the reference synthetic image from a simulated hologram, aiming to minimize the differences and identify optimal parameters. At step 2, we take the raw simulated hologram as the input for the Holo-processing with the identified optimal parameters in step 1, which helps to convert the simulated hologram to better synthetic images. These better synthetic images are the input images for the training of the U-Net.

In summary, we propose a LBS method to generate better, highly realistic data by utilizing a single experimentally captured

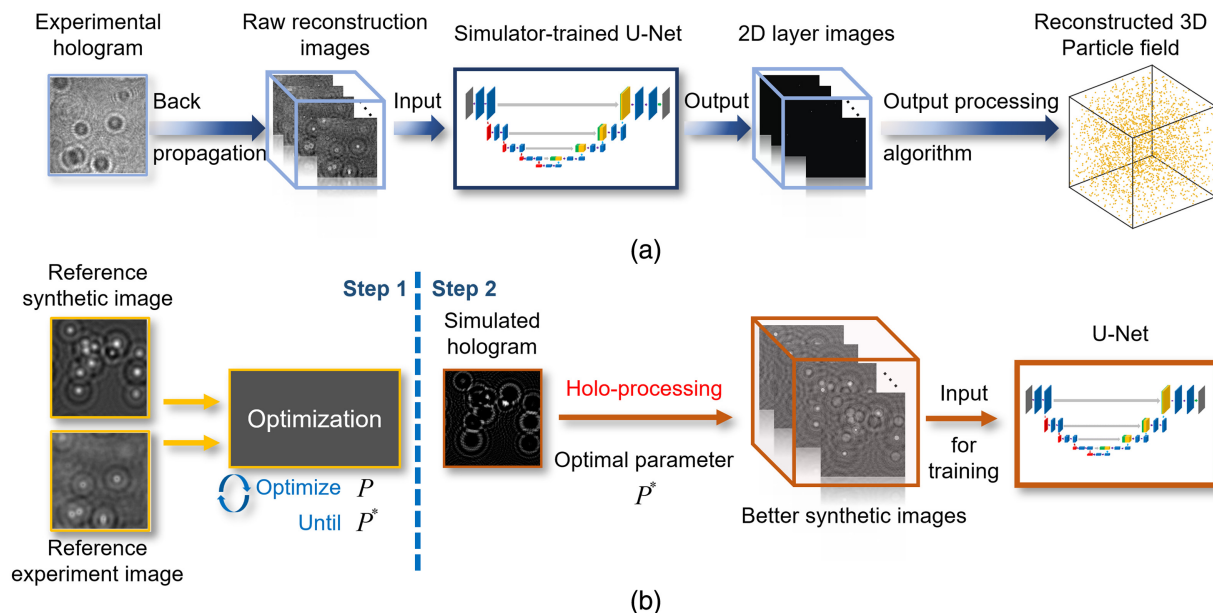


Fig. 1 LBS method framework. (a) The reconstruction process for the experimental 3D particle field. (b) The training data preparation for the U-Net. P is the parameter to be optimized, and P^* is the optimal parameter.

image sample and through an optimization method, without the help of any neural networks. In addition, we validate the quality of the generated better synthetic data with a simple U-Net. We believe that this technique not only paves the way toward the application of 3D particle field monitoring but also holds the potential in addressing complex problems that either lack a ground truth or in situations where obtaining a ground truth is significantly prohibitive. This is particularly applicable in scenarios involving multiple-scattering imaging, such as speckle-denoising,³³ imaging in dynamic scattering media,³⁴ and the phase recovery of medical biological tissues.³⁵ Our LBS method opens up a new paradigm in creating highly realistic synthetic data for deep learning-based imaging techniques.

2 LBS Method

In this section, we introduce the framework of the LBS method in detail. The LBS method includes high-contrast simulation (HCS), optimization in combination with experiments (OCE), and a modified U-Net model.

2.1 HCS Framework

Figure 2 shows the details of the HCS method. The HCS intends to generate simulation images with high contrast to train the modified U-Net. To be more specific, we want to obtain the simulated hologram that is to be used for the generation of training input data next and the corresponding ground truth images as labels.

We prepare a stack of layered images with a size of $256 \text{ pixel} \times 256 \text{ pixel}$, inside which are a certain number of particles (the number can be 0, that is, no particles). In the simulated layer images, the gray value of the pixels occupied by particles is 1 (white), whereas the gray value of the background area is 0 (see sec. 1A in the [Supplementary Material](#)). These layered images are equally spaced along the optical axis (z -axis), forming a 3D particle field. Each layer image is illuminated by an on-axis plane wave and diffracts onto the sensor plane using the angular spectrum (AS) diffraction model.³⁶ All the complex diffraction distributions are superimposed, and the intensity of the synthetic distribution is the original hologram (see sec. 1B in the [Supplementary Material](#)). The DC component is then subtracted from the hologram. The simulated hologram is preserved in its continuous gray-scale form instead of being quantized into gray-level steps to simulate the hologram recorded by an image sensor. The reason behind this is that, based on our observation, the quantization process of

the hologram introduces a certain level of distortion to the frequency spectrum of the simulated hologram. This distortion then affects the subsequent image reconstruction performance through backpropagation, resulting in less ideal outcomes (see sec. 1C in the [Supplementary Material](#)).

To make the output processing algorithm [Fig. 1(a)] more efficient, we encode the layer images into ground truth images as the training labels of the modified U-Net. Specifically, in ground truth images, each particle corresponds to a square pattern measuring $3 \text{ pixel} \times 3 \text{ pixel}$. For a particle with a diameter of D_j on a layer (if present), its central position (x_j, y_j) coincides with the central position of the pattern in the ground truth image. The gray value of the 9 pixels in the pattern is equal to $D_j \times 4$. For instance, a particle with a diameter of $25 \mu\text{m}$ corresponds to a gray value of 100. These ground truth images will later be used as training labels for a U-Net network.

2.2 OCE Framework

The simulated hologram generated by the HCS method is then backpropagated to different distances to generate the raw reconstruction images (see sec. 1B in the [Supplementary Material](#)). If we aim to generalize a simulator-trained model into the experiment, these ideally simulated raw reconstruction images without any noise introduced by the experimental setup and process are not suitable for the training of the modified U-Net. There is a lot of noise in the experiments, and even in the same experiment, sometimes we need to adjust the exposure time of the image sensor or the illuminating light intensity on the objects, which may lead to dramatic changes in the background light intensity in the captured hologram. Therefore, the experimentally raw reconstructed images are generated by holograms with different noises and different DC terms. To make simulated data similar to experimental data, the OCE framework facilitates the incorporation of experimentally introduced noise into the simulated hologram and subtracts the influence of fluctuating illuminated light intensity. The averaging of all raw reconstructed images acquired from a 3D particle field effectively removes random noise, thereby directly revealing the system noise introduced by the experiment and the background light intensity. We use the Gaussian noise and Gamma transformation to synthesize the system noise introduced by the experiment and by the average light intensity, respectively. We identify the optimized Gaussian noise parameters and the Gamma transformation factor by comparing the average of raw reconstructed images obtained from an experimental hologram with those obtained from a simulated hologram. By utilizing the optimal parameters, we make the reconstruction images obtained by simulation more similar to those obtained by the specific experiment. The OCE framework includes data preparation, a Holo-processing module, and an Optimization module for the optimization of the noise parameters and Gamma transformation parameter, which is illustrated in Fig. 3.

2.2.1 Data preparation and Holo-processing

As shown in Fig. 3(a), we used an in-line holography setup to capture particle field images in the experiment. The captured experimental hologram is cropped to a size of 256×256 , which is a representative region of the entire experimental conditions, and generally meets the requirement that the image quality should not be poor and there should not be too many impurities inside. It is then backpropagated to get a series of raw

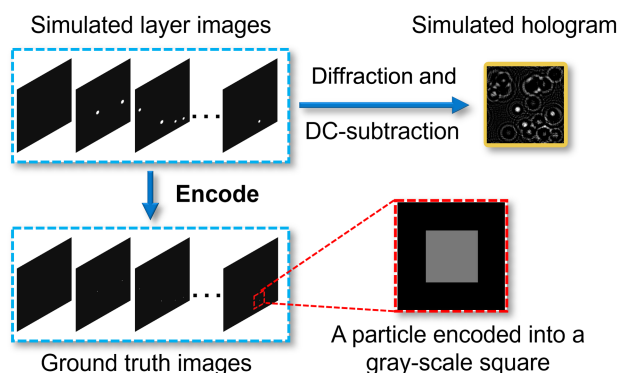


Fig. 2 HCS framework.

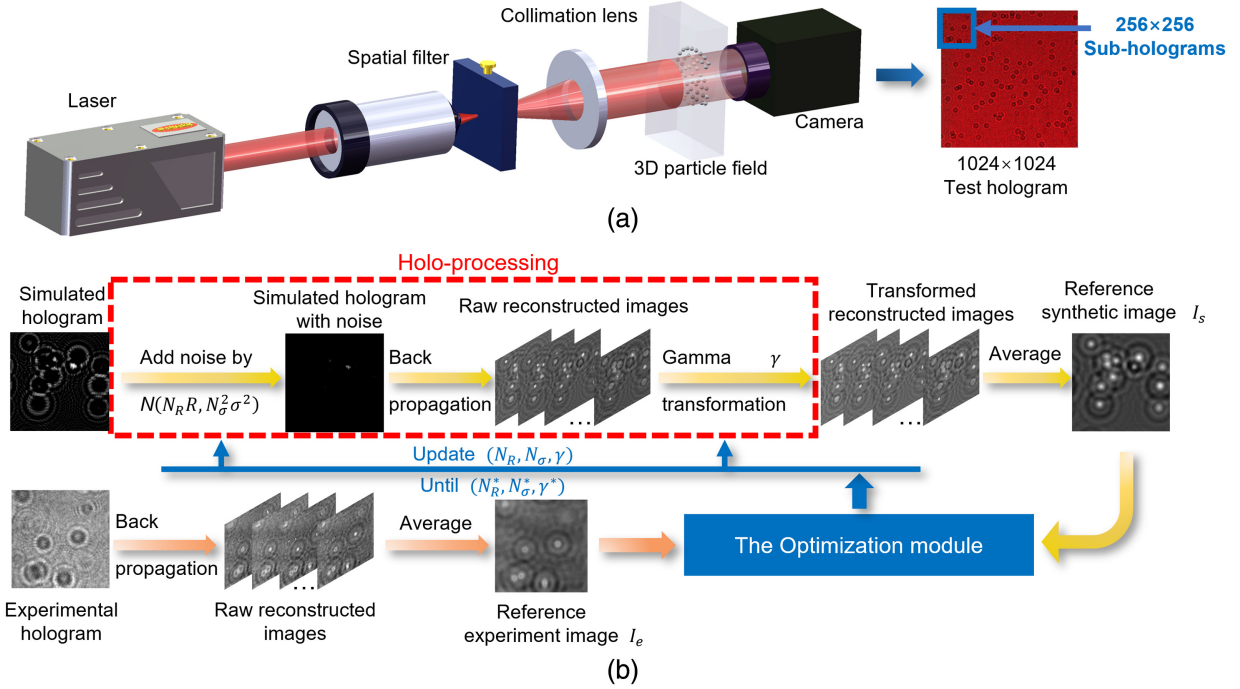


Fig. 3 (a) In-line holography experimental setup and data preparation. (b) OCE framework.

reconstructed images I_i , and then, we average these images to get I_e ,

$$I_e = \frac{1}{n} \sum_{i=1}^n I_i, \quad (1)$$

where I_e is the average of raw reconstructed images by experiment, I_i is the i th raw reconstructed image, and n is the number of reconstruction images. As shown in Fig. 3(b), we take this average image as the reference experimental image as the input for the subsequent optimization module in the OCE method. On the other hand, in the simulation, as in Sec. 2.1, we model the 3D particle field by layer images and obtain the simulated hologram. Now, we propose the Holo-processing, which aims to generate better reconstructed images so that we can average them to get the synthetic average image as the other input for the subsequent optimization module in the OCE method.

In Holo-processing, first, we add Gaussian random noise to the raw simulated hologram. However, as we explained in Sec. 2.1, we did not quantize the hologram to the image format, so strictly speaking, our hologram does not exist in the form of image data. Because of that, we cannot specify the value of the mean and variance of Gaussian noise and then add it directly to the hologram as others do. Instead, initially, we calculate the range R (the difference between the maximum and minimum values) and the standard deviation σ of the raw simulated hologram. Then, we specify the appropriate multiples N_R and N_σ to control the level of added noise. Random numbers of size 256×256 are generated with a normal distribution obeying $N(N_R \cdot R, N_\sigma^2 \cdot \sigma^2)$. By adding noise at this level to the simulated hologram, we get a new hologram. The new hologram with updated noise is then backpropagated to create raw reconstructed images. After that, we apply image enhancement to

all the raw reconstructed images using gamma transformation,³⁷ which is expressed as

$$s = cr^\gamma, \quad (2)$$

where r is the gray value of input, s is the gray value of output, c is the gray scale factor and it is equal to 1 here, and γ is the Gamma factor, which controls the scale of the entire transformation. These transformed reconstructed images after gamma transformation are the output of the Holo-processing module. We still average them to obtain the reference synthetic image I_s . Subsequently, the reference experimental image I_e and the reference synthetic image I_s are sent into the optimization module together as input.

2.2.2 Optimization

The objective of the Optimization module is to determine the optimal values for three parameters (N_R, N_σ, γ) such that the input reference synthetic image closely matches the reference experimental image. To accomplish this, we aim to solve the following optimization problem:

$$(N_R^*, N_\sigma^*, \gamma^*) = \arg \min_{N_R, N_\sigma, \gamma} 1 - \text{SSIM}[I_e, I_s(N_R, N_\sigma, \gamma)], \quad (3)$$

where $(N_R^*, N_\sigma^*, \gamma^*)$ are the corresponding optimal parameters for (N_R, N_σ, γ) , I_e and I_s are the reference experimental image and reference synthetic image, respectively, and SSIM is the structural similarity index measure.³⁸

The problem in Eq. (3) is a complex, high-dimensional, non-convex nonlinear optimization problem. It is challenging to formulate a precise mathematical expression for this problem, let alone obtain complete gradient information. Moreover, empirical investigations using gradient-based optimization methods

have revealed many local optimal solutions, causing the algorithms to become easily trapped and unable to escape from these local optima. To address this, we employ the genetic algorithm (GA),^{39,40} which is a highly effective global optimization method for solving intricate high-dimensional problems, especially for problems that are not smooth, not derivable, or have no explicit mathematical model (see sec. 1F in the [Supplementary Material](#)).

The optimal parameters (N_R^* , N_σ^* , γ^*) obtained from the optimization module are utilized as inputs to the Holo-processing module. Specifically, the noise distribution determined by N_R^* and N_σ^* is incorporated into the simulated hologram. Subsequently, the raw reconstructed images are derived, and a gamma transformation using γ^* is applied to them. The resulting better synthetic images, which best replicate the experimental noise, together with the corresponding ground truth images by the HCS framework, are utilized as inputs and labels for training the modified U-Net model.

2.3 Modified U-Net

U-Net is a network architecture widely adopted in image segmentation tasks, particularly for medical images⁴¹ or images related to scattering media,²² where spatial precision is critical. It is especially well-suited for pixel-wise prediction tasks as it generates outputs that are spatially aligned with the inputs, enabling dense, per-pixel mapping.

We have made some modifications based on the original U-Net,⁴¹ as shown in Fig. 4. We use mean square error as the loss function and choose Adam optimizer as the optimizer, and the learning rate is set to 10^{-4} or 10^{-5} according to specific conditions. When testing the network, the original predictive output image must pass through the image processing algorithm (see sec. 1D in the [Supplementary Material](#)) to obtain the required particle field information. After that, we apply the evaluation metrics to quantify the result (see [Appendix](#) for more details).

The U-Net architecture is implemented using Keras 2.6.0 with the TensorFlow-GPU 2.6.0 backend. Other software environments are Python 3.7, CUDA 11.2, CUDNN 8.1, and MATLAB R2022b. The network training is conducted on an

Nvidia RTX 3090 24G GPU, and the generation of training data is conducted on an Intel Core i5-11400H CPU.

3 Results

3.1 Simulation Results

In this section, we show how far we can go in the limit of particle concentration on the simulation side. We choose the particle per pixel (ppp) concentration metric^{21,22} here (see sec. 2A in the [Supplementary Material](#)).

We model a 3D particle field, with 60 particles with diameters of 20 to 30 μm , 63 layers with an axial layer interval of 30 μm , and an axial distance range of 1800 to 3660 μm . We take 26.82 μm as the lateral distance of a voxel. It is worth pointing out that, in the simulation, the transformed reconstructed images are obtained by Holo-processing with the original parameters (0, 0, 1), given the fact that there is no experimental data involved here. Because each hologram contains 256 pixel $\times$ 256 pixel with a pixel pitch of 2.2 μm , each group of data thus corresponds to a particle field of 563 $\mu\text{m} \times 563 \mu\text{m} \times 1860 \mu\text{m}$ (21 voxel $\times$ 21 voxel $\times$ 63 voxel). A total of 128 groups of training data are prepared, corresponding to a data shape of 256 $\times$ 256 $\times$ 8064, and the 100 epochs of training take 135 min with a learning rate of 10^{-4} .

In the test, we prepared nine groups of data and spliced them in a horizontal 3 $\times$ 3 way, so the particle field size after splicing is 63 voxel $\times$ 63 voxel $\times$ 63 voxel. As each group of data contains 60 particles, this particle field contains 540 particles; that is to say, the ppp concentration is $\sim 1.36 \times 10^{-1}$. Compared with 6.1×10^{-2} , the highest concentration used by Shao et al.,²² this concentration is very high. We show how difficult it is to identify particles at this high concentration in Fig. 5(a), where two transformed reconstructed images at different distances [defined by Fig. 3(b)] are shown. Perhaps A1 and A4, the two in-focus particles, can be identified using the traditional image threshold method or iterative calculation method, but it is impossible to identify A2 and A3 using those methods. A2 and A3 are seriously disturbed by the defocused image of other particles, and even partially occluded, so only part of the particle profiles is exhibited. For such cases, where the particles overlap in the

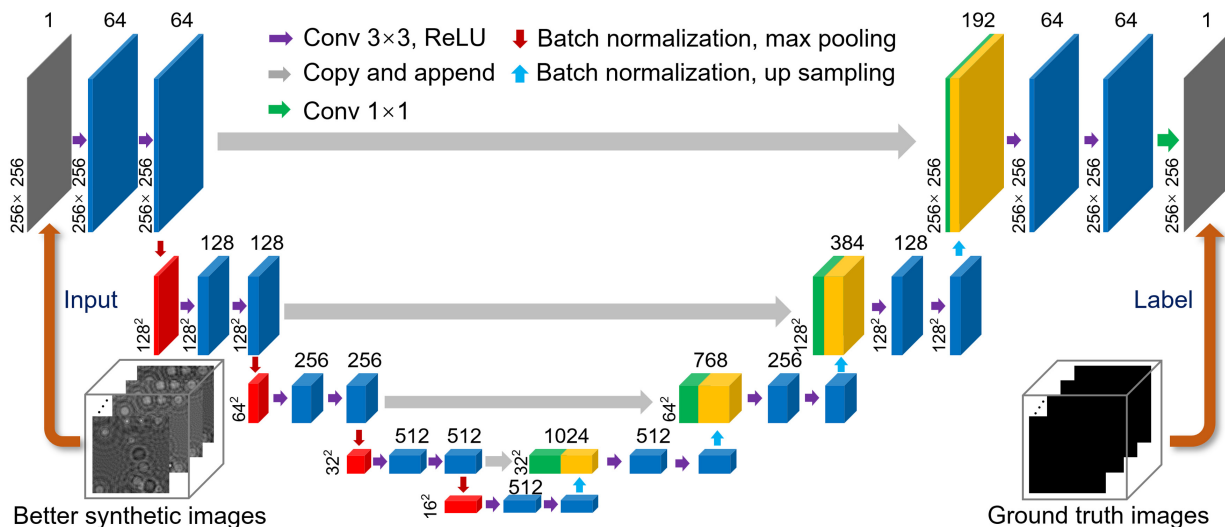


Fig. 4 Modified U-Net framework with the training input and label.

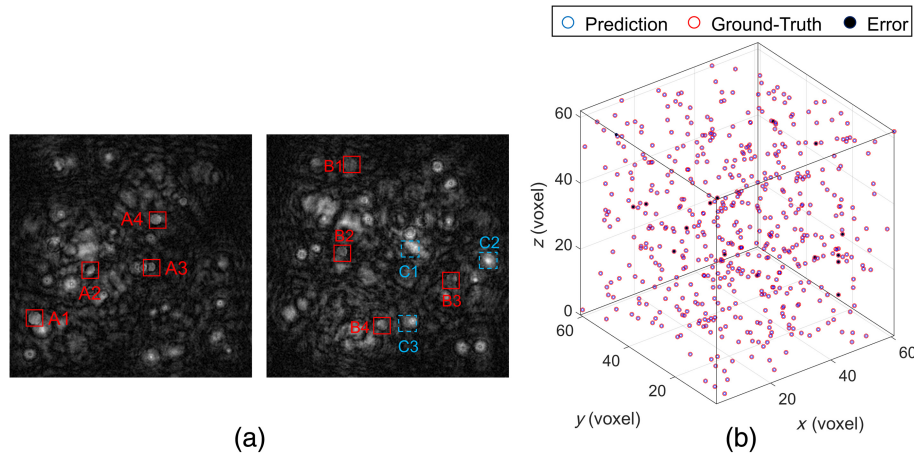


Fig. 5 (a) Two transformed reconstructed images with additional name labels, with the ppp concentration is 1.36×10^{-1} . The solid red box marks all the in-focus particles and the dashed blue box marks several out-of-focus particles. (b) Location comparison between the prediction result and the ground-truth.

axial direction, it is difficult to obtain accurate results using traditional methods. In addition, the traditional methods are generally based on light intensity (gray level), so the four in-focus particles B1 to B4 are likely to be considered the background because of their similar gray values compared with the surrounding environment, whereas the three out-of-focus particles C1 to C3 are likely to be considered in-focus particles because of their larger brightness. Figure 5(a) fully demonstrates the difficulty and importance of reconstructing high-concentration particle fields.

Here, like others, both the lateral positioning error and the axial positioning error should be less than 1 voxel ($26.82 \mu\text{m}$ in transverse direction, $30 \mu\text{m}$ in axial direction) to be considered a correct location of a particle. Figure 5(b) demonstrates the reconstructed particle field obtained using the reconstruction method proposed in this paper under the simulation conditions described in this section. The reconstructed particle field and the simulated ground truth particle field are displayed simultaneously in the same coordinate system, with particles exhibiting positional errors marked in black. Among these more than 500 particles, the number of particles of false negative (FN) is 15 and the number of particles of false positive (FP) is 1 (see Appendix for related definitions). All 16 particles are considered error particles. The result shows that even at such a high concentration (2.2 times 6.1×10^{-2} ppp), we still obtain the extraction rate (ER) of 97.21% and Jaccard index (JI) reached 0.97, which is better than both Shao et al.'s²² ER of 95% at 6.1×10^{-2} ppp and Chen et al.'s²¹ ER of 82.4% at 6.1×10^{-2} ppp. The results in Fig. 5(b) show that, after using the proposed method, the reconstruction result accuracy reaches a very high level under high concentration.

3.2 Experimental Results

We use polystyrene microspheres with diameters of 20 to $30 \mu\text{m}$ freely suspended in ultrapure water, which is placed in a quartz cuvette with a thickness of 3 mm, as the particle field for the experiment (see sec. 2B in the Supplementary Material). The particle concentration measures 10,000 particles/mL when the solution is diluted. A hologram is captured by the image sensor, and a region with a size of 1024×1024 in the whole

hologram is cropped and selected as the test data [Fig. 3(a)]. We crop this region to 4×4 sub-holograms with a size of 256×256 and backpropagate them to get 976 reconstructed images (1800 to $4800 \mu\text{m}$, interval $50 \mu\text{m}$, shape $256 \times 256 \times 976$). It is important to note that to align experimental images with synthetic ones, where particles appear white instead of black (low gray value), we process the experimental reconstructed images using image inversion ($255 - I_o$) on the original backpropagated images. To examine the performance of the test result, we use the self-made compound focusing criterion (see sec. 1E in the Supplementary Material) and manual examination to obtain the reference results, which are used to evaluate the results by the output of the simulator-trained model. These reference results are not the ground truth for the training of the U-Net but for the evaluation of the prediction by the simulator-trained U-Net. When validating the network, these raw reconstructed experimental images are taken as the input. Through the simulator-trained U-Net and the output processing algorithm, the prediction results are generated and then evaluated by comparing them to the reference results. As the pixel size of the image sensor is $2.2 \mu\text{m}$, the volume size corresponding to the test image is $2253 \mu\text{m} \times 2253 \mu\text{m} \times 3000 \mu\text{m}$.

To generate better synthetic training data, we must find the optimal parameters for the Holo-processing first. On the simulation side [Fig. 3(b)], as presented in Sec. 2.1, we generate the random particle field containing 12 particles with diameters of 20 to $30 \mu\text{m}$. The z -direction distance from the layers in the particle field to the sensor plane is 1800 to $4800 \mu\text{m}$, with its interlayer interval to be $200 \mu\text{m}$. One simulated hologram is generated and used as the reference simulated hologram. On the experiment side, among the 16 subholograms mentioned above, we choose one with an image quality that is relatively good without much impurity, to be the experimental hologram for reference. During optimization, by setting the fitness function as the objective function in Eq. (3), the GA is optimized for up to 100 generations, unless termination conditions are met. We conduct multiple runs of the GA within a time duration of 20 min. Although the optimal parameters (N_R^* , N_σ^* , γ^*) varied with each run, they are found to be close to each other. We select the set of parameters that minimizes the fitness function as the

best parameter values, which in this case are $(-0.3068, 0.2164, 0.8337)$. After that, we directly send $(-0.3068, 0.2164, 0.8337)$ into Holo-processing as parameters and start to generate all the 256 groups of data processed by Holo-processing (1800 to $4800\ \mu\text{m}$, interval $200\ \mu\text{m}$, 12 particles per group). Then, we get 4096 better synthetic images (shape $256 \times 256 \times 4096$) as the training input images. Figure 6 shows the advantage of the better synthetic data over normal simulation data, which is not processed by the HCS method or the OCE method.

As observed, of the three defocused particles, namely, A1, A2, and A3, A2 virtually mirrors A3, whereas A1 diverts noticeably by featuring an unusual black core. This attribute might become a significant challenge for the model when attempting to distinguish between in-focus particles. On a similar note, the in-focus particle B2 bears a striking resemblance to B3, whereas B1 contrasts starkly, being partially obscured by surrounding noise. Moreover, the in-focus particles C2 and C3 share strong similarities, yet C1 is disproportionately drowned out by the background. Hence, the attributes of the better synthetic data align more closely with the characteristics of actual experimental data.

With the corresponding 4096 synthetic ground truth images from the simulated particle fields by HCS as the labels, the U-Net model is trained for 100 epochs with a learning rate of 10^{-5} , which takes ~ 80 min. During training and validation, we consider particles with lateral errors within 10 pixels ($22\ \mu\text{m}$) and axial errors within $150\ \mu\text{m}$ to be true positive (TP). To evaluate the overall performance, we set another indicator point, which is equal to $(\text{ER} + \text{JI})$ when the test particle number is over a certain number, or 0 when that number is below this certain number (see sec. 2B in the [Supplementary Material](#)).

Figure 7 shows the results of our training and validation. As a comparison, we also used the other two types of data from Fig. 6

for training. Specifically, for the normal simulation data, we used the exact same particle field as in the better synthetic data, meaning the labels are identical between the two, with the only difference being the training input, whereas all other parameters remain consistent. For the real experiment data, we used another hologram captured at a different time to create the training input. As for the labels corresponding to these training inputs, we directly employed the focusing method mentioned in Ref. 42 to locate the in-focus positions of the particles to generate the labels (though they are not fully accurate), whereas all other parameters remain consistent. We can see that for better synthetic data, after only 100 epochs of training, with a training time of less than 80 min, at the 84th epoch, we get a good result, that is, JI reached 0.886 and ER reached 0.8557 (85.57%). However, for the normal simulation data, the highest JI is only around 0.3, and the ER does not exceed 30%. On the other hand, for the real experimental data, the highest JI is below 0.7, and the ER reaches only $\sim 50\%$. In addition, the model trained with better synthetic data is the most stable during validation. In contrast, the model trained with real experimental data is highly unstable, and its training loss does not decrease to a sufficiently low value. Meanwhile, the performance of the model trained with normal simulation data on the validation dataset shows no significant improvement. This result indicates that although using real experimental data as input is better than using simulation data, this is only true if the corresponding labels are sufficiently accurate. When we cannot ensure accurate labels, using simulation data with perfectly accurate labels may be a better option. Furthermore, when comparing the two types of simulation data—better synthetic data and normal simulation data—it becomes clear that the realism of the simulation data is quite crucial to the network's performance.

After the training, using the trained models, we can get the output prediction layer images. Figure 8 shows the comparison

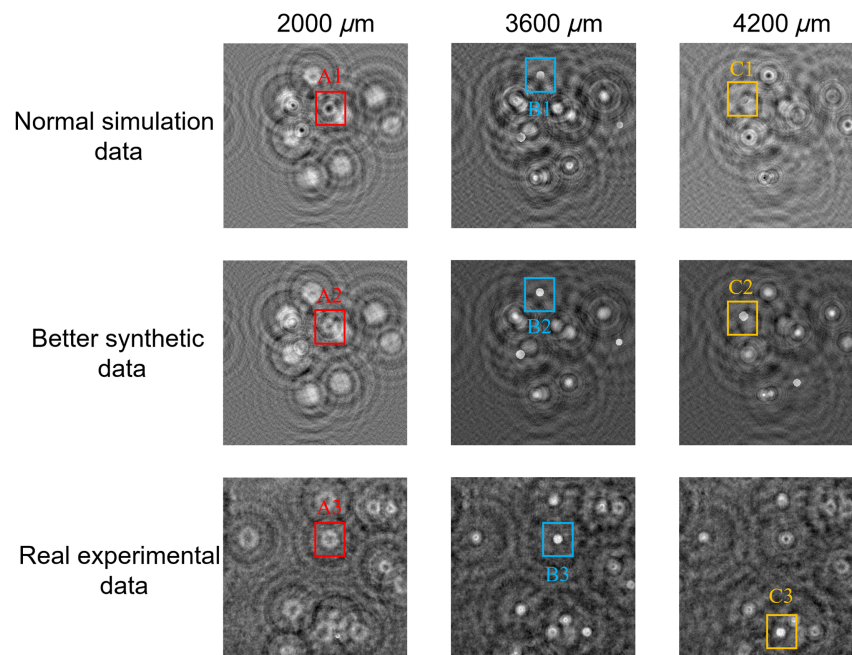


Fig. 6 Reconstruction comparison among the normal simulation data, better synthetic data, and real experimental data under different reconstruction distances.

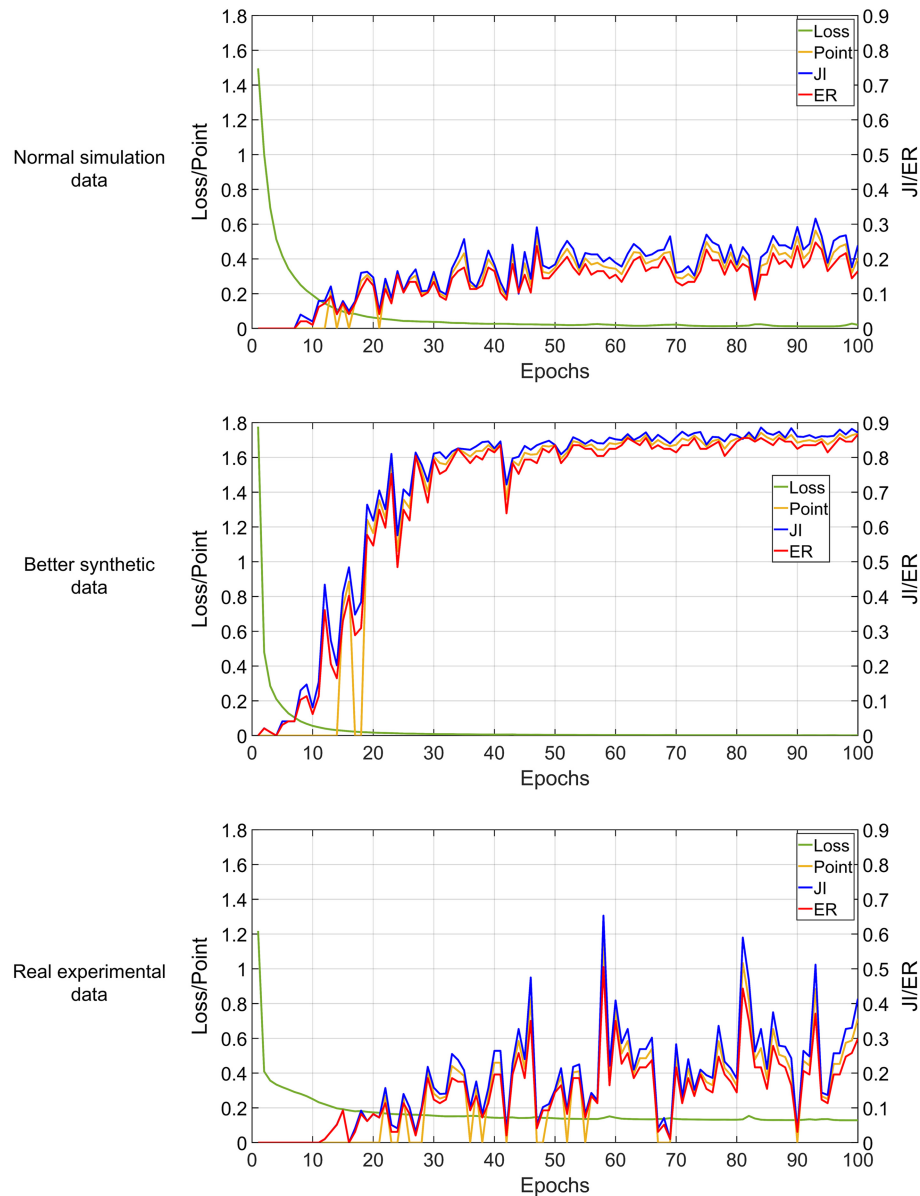


Fig. 7 Training results for the three types of training datasets: loss during training, and JI, ER, and point metrics for validation.

between the predicted experimental results and the reference results. Here, the experimental results are generated by three U-Nets trained on three types of data, respectively: normal simulation data, better simulated data (obtained through the OCE framework proposed in this paper), and real experimental data. The particles of TP, FN, and FP are marked with different colors in the figure. We can see that, for our better synthetic data result, in the lateral direction, most of the reconstructed particles match the reference result very well, and in the axial direction, although some particles do not match the reference result exactly, they are still within our tolerance ($150\ \mu\text{m}$). Moreover, by inspecting these FN particles' positions, we find that most of them are at the edge of the sub-images with a size of $256\ \text{pixel} \times 256\ \text{pixel}$. In other words, the primary reason the model fails to detect some particles is that crucial information at the edges is lost during the image cropping process. Therefore, these undetected FN particles are reasonable and can be

accepted, which, to some extent, also shows the superiority of our LBS method. On the other hand, in the reconstruction from the model trained with normal simulation data, most of the particles are missed. Although the reconstruction from the model trained with real experimental data performed better than the normal simulation data, it is clearly much worse compared with the results from the better synthetic data. The results in Fig. 8 fully demonstrate the effectiveness of the proposed method.

For comparison, the concentration in Chen et al.'s²¹ experimental validation is around 50 particles/mL, and its ER is similar to ours, but its axial error is $\sim 1\ \text{mm}$, which is much larger than ours ($150\ \mu\text{m}$). In addition, compared with Tahir et al.'s³² experimental result, our requirement for TP is stricter than theirs, which is a 4 particle size error in the transverse direction and a 12 particle size error in the axial direction. In addition, we did not encounter the problem that appeared in their experiments, where the performance drops dramatically when the depth increases.

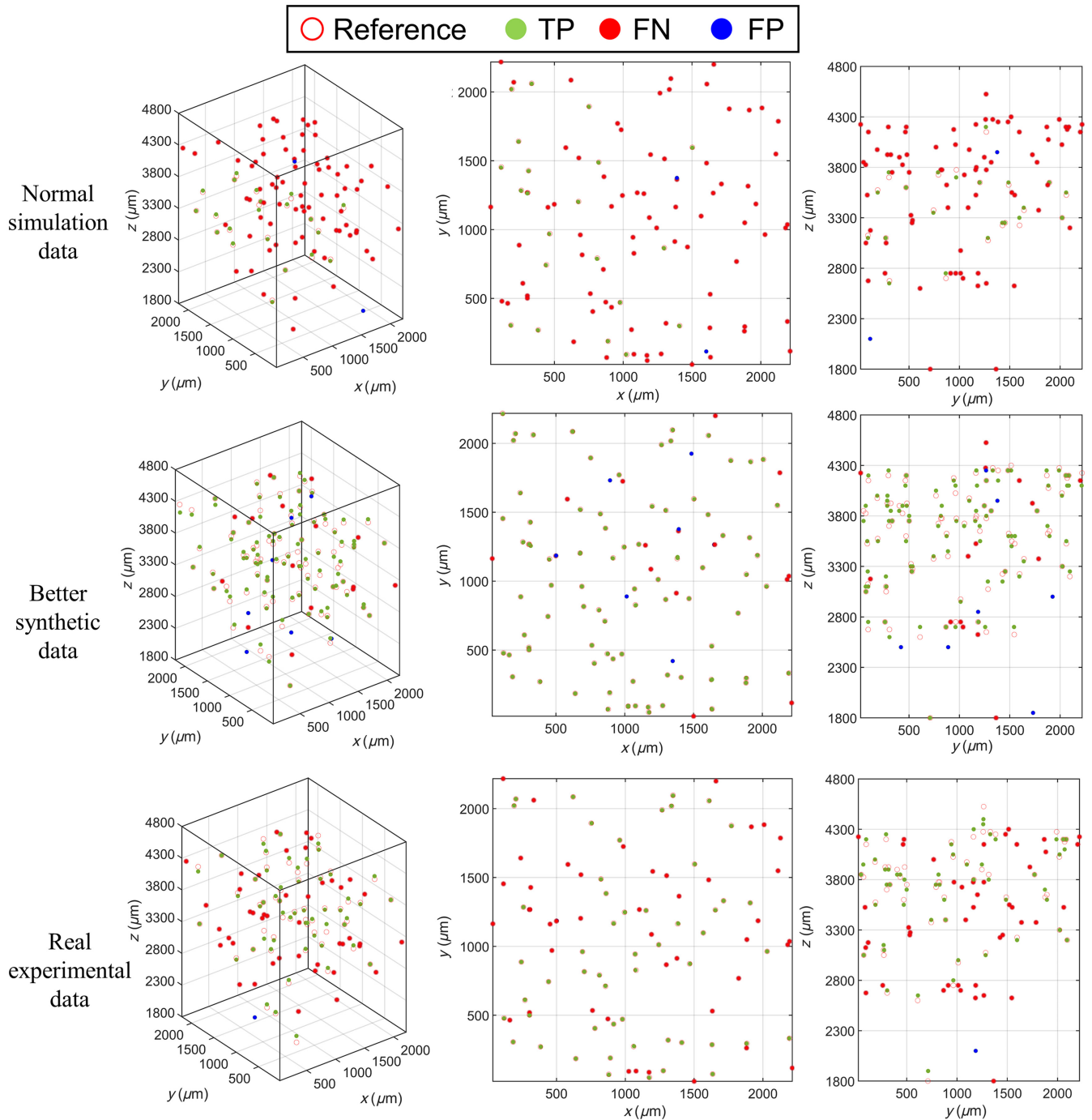


Fig. 8 Comparison of particle locations between the reconstruction and the reference results for the three types of training datasets. The reconstructed particles are categorized as true positive (TP), false negative (FN), and false positive (FP).

As we can see from Fig. 8, even when the depth increases to over 4000 μm , our performance is still good.

4 Discussion and Conclusion

In this paper, we present and experimentally demonstrate a fast and adaptive method called the LBS method for generating highly realistic synthetic images for scattering media. Without any help from a neural network, after an optimization process

is used to best synthesize the noise caused by multiple scattering, defocusing, the experimental system, and environment from a single measurement and added to the simulated data, the synthetic data are obtained. To validate the quality of the better synthetic data, we use it to train a simple U-Net model. We demonstrate the effectiveness of the created highly realistic synthetic data and the generalization capability of the simulator-trained U-Net model by the accurate reconstruction of real high-concentration 3D particle fields. The major challenge in reconstructing a 3D particle

field with high concentration is a complex multiple scattering process.³¹ Designing a network specifically tailored to this process can be exceedingly complex, and the training cost can be tremendous.^{21,32} However, in this paper, we successfully generate synthetic data that accurately reflect the impact of this complex physical process. Therefore, the proposed LBS method perfectly addresses the two critical issues in 3D particle field monitoring, which are the absence of experimental ground truth and the complex physical model.

The new concept that we propose in this paper, which is to synthesize the complex physical model output by optimizing the synthetic data using experimental output, offers inspiration and guidance to others employing simulation-based training methods, where paired ground-truth labels are also hard to obtain. For instance, medical imaging of tissue samples (phase objects) has predominantly depended on time-intensive staining procedures due to the significant barrier presented by multiple scattering in deep biological tissue. The implementation of the proposed HCS and OCE methods could potentially enable us to swiftly generate a sizable dataset with corresponding high-contrast ground truth images. This would negate the need for laborious processes such as mechanical movements and repeated microscopic imaging to acquire training data.

This strategy of simulator-trained deep learning will make significant contributions to the next-generation deep learning methods,^{26,43} where strenuous manual labeling is no longer needed. We expect that the successful creation of synthetic data adhering to the multiple-scattering process will amplify our comprehension of this process in complex scattering media,⁴⁴ potentially sparking groundbreaking formulations or models in this field. In addition, we expect that this LBS method will enhance the applicability of three-dimensional particle field imaging in areas such as particle tracking velocimetry,^{45,46} holographic cytometry,^{47–50} and water quality measurement.^{51–53}

Despite the promising results, our approach has certain limitations. Specifically, we have validated the method only on polystyrene microspheres under a limited range of concentrations, which may impose an upper bound on the sample density. Although our framework does not require explicit calibration of the experimental system—as the LBS method inherently captures the system’s noise characteristics and illumination conditions—more complex, denser, or thicker scattering media may require further modeling of the physical interactions, including more sophisticated noise, transformation representations. Although the noise characteristics are system-specific, our method is general and can be applied to other systems, such as in-line holographic particle imaging, if researchers choose to adopt our approach.

Our method can be further improved in the future. Notably, the optimal SSIM value between the reference synthetic and experimental images currently stands at ~ 0.7 . If we can increase this value closer to 1, thereby enhancing the similarity between the synthetic and experimental images, it is likely that the network’s performance could also be improved. One promising approach is using a generative adversarial network.⁵⁴ This technique, given a training set, learns to produce new data that statistically resembles the training set data. This method holds potential for generating superior synthetic data similar to experimental data, which could improve the model’s generalization and broaden its application across various research fields.

5 Appendix: Evaluation Metrics

We typically focus on two values, namely, extraction rate and accuracy. In the result of network prediction, there are three kinds of particles, namely, true positive (TP), false positive (FP), and false negative (FN). TP represents correctly positioned particles, FP represents error-detected particles that are in the reconstruction but not in the ground truth, and FN represents missing particles that are not detected in the reconstruction but are present in the ground truth. The extraction rate (ER) is defined as

$$ER = \left(1 - \frac{n_{FN}}{n_{GT}}\right) \times 100\%, \quad (4)$$

where n_{FN} is the number of particles of FN and n_{GT} is the number of particles in the ground truth or in the reference result. Specifically, the accuracy metric we use here is the Jaccard index (JI), which is computed as

$$JI = \frac{n_{TP}}{n_{TP} + n_{FP} + n_{FN}}. \quad (5)$$

Noting that the predictions of our network cannot be the same as the ground truth, we set a certain tolerance when judging whether a particle is TP and the tolerance is different in simulation and experiment.

Disclosures

The authors declare no conflict of interest.

Code and Data Availability

All the data and methods needed to evaluate the conclusions of this work are presented in the main text and the [Supplementary Material](#). Additional data are available from the corresponding author upon reasonable request.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant Nos. 2022YFB2804300 and 2023YFB3611500), the National Natural Science Foundation of China (Grant No. 62275218), the China Postdoctoral Science Foundation (Grant No. 2022M712586), and the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023A1515011335).

References

1. X. Li et al., “Polarimetric imaging through scattering media: a review,” *Front. Phys.* **10**, 815296 (2022).
2. X. Li et al., “Polarimetric image recovery method combining histogram stretching for underwater imaging,” *Sci. Rep.* **8**, 12430 (2018).
3. L. Gong, S. Lin, and Z. Huang, “Stimulated Raman scattering tomography enables label-free volumetric deep tissue imaging,” *Laser Photonics Rev.* **15**, 2100069 (2021).
4. H. Zhuang et al., “High speed color imaging through scattering media with a large field of view,” *Sci. Rep.* **6**, 32696 (2016).
5. D. Huang et al., “Optical coherence tomography,” *Science* **254**, 1178–1181 (1991).
6. C.-Y. Shen et al., “Multiplane quantitative phase imaging using a wavelength-multiplexed diffractive optical processor,” *Adv. Photonics* **6**, 056003 (2024).

7. T. Kim et al., "Solving inverse scattering problems in biological samples by quantitative phase imaging," *Laser Photonics Rev.* **10**, 13–39 (2016).
8. S. Xu et al., "Tensorial tomographic Fourier ptychography with applications to muscle tissue imaging," *Adv. Photonics* **6**, 026004 (2024).
9. S. Yoon et al., "Deep optical imaging within complex scattering media," *Nat. Rev. Phys.* **2**, 141–158 (2020).
10. M. Lyu et al., "Learning-based lensless imaging through optically thick scattering media," *Adv. Photonics* **1**, 036002 (2019).
11. F. Wang et al., "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).
12. Y. Xue et al., "Reliable deep-learning-based phase imaging with uncertainty quantification," *Optica* **6**, 618–629 (2019).
13. L. Huang et al., "Holographic image reconstruction with phase recovery and autofocusing using recurrent neural networks," *ACS Photonics* **8**, 1763–1774 (2021).
14. H. Wang et al., "Deep learning enables cross-modality super-resolution in fluorescence microscopy," *Nat. Methods* **16**, 103–110 (2019).
15. T. Liu et al., "Deep learning-based super-resolution in coherent imaging systems," *Sci. Rep.* **9**, 3926 (2019).
16. Y. Li, Y. Xue, and L. Tian, "Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media," *Optica* **5**, 1181–1190 (2018).
17. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**, 921–943 (2019).
18. K. Wang et al., "On the use of deep learning for phase recovery," *Light Sci. Appl.* **13**, 4 (2024).
19. X. Hu et al., "Metasurface-based computational imaging: a review," *Adv. Photonics* **6**, 014002 (2024).
20. Y. Rivenson et al., "Phase recovery and holographic image reconstruction using deep learning in neural networks," *Light Sci. Appl.* **7**, 17141–17141 (2018).
21. N. Chen, C. Wang, and W. Heidrich, "Holographic 3D particle imaging with model-based deep network," *IEEE Trans. Comput. Imaging* **7**, 288–296 (2021).
22. S. Shao et al., "Machine learning holography for 3D particle field imaging," *Opt. Express* **28**, 2987–2999 (2020).
23. Y. Wu et al., "Dense-U-net: dense encoder–decoder network for holographic imaging of 3D particle fields," *Opt. Commun.* **493**, 126970 (2021).
24. W.-N. Li et al., "Short U-net model with average pooling based on in-line digital holography for simultaneous restoration of multiple particles," *Opt. Lasers Eng.* **139**, 106449 (2021).
25. Z. Hao et al., "Characterization method for particle extraction from raw-reconstructed images using U-net," *Front. Phys.* **9**, 816158 (2022).
26. C. M. de Melo et al., "Next-generation deep learning based on simulators and synthetic data," *Trends Cognit. Sci.* **26**, 174–187 (2022).
27. S. I. Nikolenko, *Synthetic Data for Deep Learning*, Springer (2021).
28. G. E. Granados et al., "Towards a multi-fidelity deep learning framework for a fast and realistic generation of ultrasonic multi-modal Total Focusing Method images in complex geometries," *NDT & E Int.* **139**, 102906 (2023).
29. G. E. Granados et al., "Generative domain-adapted adversarial auto-encoder model for enhanced ultrasonic imaging applications," *NDT & E Int.* **148**, 103234 (2024).
30. S. Kang et al., "Tracing multiple scattering trajectories for deep optical imaging in scattering media," *Nat. Commun.* **14**, 6871 (2023).
31. H. Wang et al., "Large-scale holographic particle 3D imaging with the beam propagation model," *Opt. Express* **29**, 17159–17172 (2021).
32. W. Tahir, H. Wang, and L. Tian, "Adaptive 3D descattering with a dynamic synthesis network," *Light Sci. Appl.* **11**, 42 (2022).
33. Z. Chen et al., "DN-GAN: denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images," *Biomed. Signal Process. Control* **55**, 101632 (2020).
34. D. Wang et al., "Non-invasive super-resolution imaging through dynamic scattering media," *Nat. Commun.* **12**, 3150 (2021).
35. S. Lin, L. Gong, and Z. Huang, "Super-resolution two-photon fluorescence tomography through the phase-shifted optical beatings of Bessel beams for high-resolution deeper tissue 3D imaging," *Laser Photonics Rev.* **18**, 2300634 (2024).
36. Y. Zhao et al., "Accurate calculation of computer-generated holograms using angular-spectrum layer-oriented method," *Opt. Express* **23**, 25440–25449 (2015).
37. J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Trans. Image Process.* **9**, 889–896 (2000).
38. W. Zhou et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (2004).
39. J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT Press (1992).
40. R. Hassan et al., "A comparison of particle swarm optimization and the genetic algorithm," in *46th AIAA/ASME/ASCE/AHS/ASC Struct., Struct. Dyn. and Mater. Conf.* (2012).
41. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci.* **9351**, 234–241 (2015).
42. E. Darakis et al., "Microparticle characterization using digital holography," *Chem. Eng. Sci.* **65**, 1037–1044 (2010).
43. J. Tremblay et al., "Training deep networks with synthetic data: bridging the reality gap by domain randomization," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit. Workshops*, pp. 969–977 (2018).
44. H. Jin et al., "Limiting the incident NA for efficient wavefront shaping through thin anisotropic scattering media," *Optica* **8**, 428–437 (2021).
45. N. Chen, C. Wang, and W. Heidrich, "Snapshot space–time holographic 3D particle tracking velocimetry," *Laser Photonics Rev.* **15**, 2100008 (2021).
46. H. Meng et al., "Holographic particle image velocimetry: from film to digital recording," *Meas. Sci. Technol.* **15**, 673–685 (2004).
47. M. Hejna et al., "High accuracy label-free classification of single-cell kinetic states from holographic cytometry of human melanoma cells," *Sci. Rep.* **7**, 11943 (2017).
48. F. Merola et al., "Tomographic flow cytometry by digital holography," *Light Sci. Appl.* **6**, e16241–e16241 (2016).
49. Z. Göröcs et al., "A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples," *Light Sci. Appl.* **7**, 66 (2018).
50. Z. Göröcs et al., "Label-free detection of Giardia lamblia cysts using a deep learning-enabled portable imaging flow cytometer," *Lab Chip* **20**, 4404–4412 (2020).
51. V. Bianco et al., "Identification of microplastics based on the fractal properties of their holographic fingerprint," *ACS Photonics* **8**, 2148–2157 (2021).
52. O. Mudanyali et al., "Detection of waterborne parasites using field-portable and cost-effective lensfree microscopy," *Lab Chip* **10**, 2419–2423 (2010).
53. Y. Li et al., "High-throughput microplastic assessment using polarization holographic imaging," *Sci. Rep.* **14**, 2355 (2024).
54. I. Goodfellow et al., "Generative adversarial nets," in *Adv. Neural Inf. Process. Syst.* **27** (2014).
55. D. Gabor, "A new microscopic principle," *Nature* **161**, 777–778 (1948).
56. T. Shimobaba et al., "Digital holographic particle volume reconstruction using a deep neural network," *Appl. Opt.* **58**, 1900–1906 (2019).

57. N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).
58. P. Memmolo et al., "Automatic focusing in digital holography and its application to stretched holograms," *Opt. Lett.* **36**, 1945–1947 (2011).
59. J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-8**, 679–698 (1986).

Bozhen Zhou is currently a graduate student at the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His current research interests include computational imaging, digital holography, and machine learning.

Zhitao Hao received his MS degree from the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. His work focuses on particle fields, digital holography, and computational imaging.

Zhenbo Ren received his BS degree in optoelectronic information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2011, his MS degree in instrument science and technology from the Tsinghua University, Beijing, China, in 2014, and his PhD in optical engineering from the University of Hong Kong, Hong Kong, in 2018. He is currently an associate professor with the School of Physical Science and Technology, Northwestern Polytechnical University, Xi'an, China. His research interests include computational imaging and intelligent sensing.

Edmund Y. Lam received his BS, MS, and doctor's degrees in electrical engineering from the Stanford University, Palo Alto, US. He is now a professor of electrical and electronic engineering and director of the Imaging Systems Laboratory at the University of Hong Kong, Hong Kong, China. He has broad research interests around the theme of computational optics and imaging and has published over 300 journal and conference papers. A recipient of the IBM faculty award, he is also a fellow of SPIE, Optica, the Institute of Electrical and Electronics Engineers (IEEE), the Society for Imaging Science and Technology (IS&T), and the Hong Kong Institution of Engineers (HKIE).

Jianshe Ma received his BS degree in mechanical engineering from the Beijing University of Science and Technology, Beijing, China, in 1992, his MS degree in machine manufacturing from the Kunming University of Science and Technology, Kunming, China, in 1997, and his PhD in mechatronic engineering from the Harbin Institute of Technology, Harbin, China, in 2000. He is currently an associate professor and a member of the Optical Memory National Engineering Research Center. His research interests include optical storage, semiconductor lighting, and ultraviolet communication.

Ping Su received her BS degree in measurement and control technology and instruments and her PhD in optical engineering from the Tsinghua University, Beijing, China, in 2005 and 2010, respectively. She is now an associate professor at the Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. Her research interests include optical storage, holography, and ultraviolet communication. She is currently a member of SPIE and Optica.