Holographic 3D particle reconstruction using a one-stage network

YUNPING ZHANG, YANMIN ZHU AND EDMUND Y. LAM*

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

Abstract: Volumetric reconstruction of a three-dimensional (3D) particle field with high resolution and low latency is an ambitious and valuable task. As a compact and high-throughput imaging system, digital holography (DH) encodes the 3D information of a particle volume into a two-dimensional (2D) interference pattern. In this work, we propose a one-stage network (OSNet) for 3D particle volumetric reconstruction. Specifically, by a single feed-forward process, OSNet can retrieve the 3D coordinates of the particles directly from the holograms without a high-fidelity image reconstruction at each depth slice. Evaluation results from both synthetic and experimental data confirm the feasibility and robustness of our method under different particle concentrations and noise levels in terms of detection rate and position accuracy, with improved processing speed. The additional applications of 3D particle tracking are also investigated, facilitating the analysis of the dynamic displacements and motions for micro-objects or cells. It can be further extended to various types of computational imaging problems sharing similar traits.

1. Introduction

Recent advances in sensor development and three-dimensional (3D) imaging techniques have attracted significant interest in 3D particle volumetric reconstruction, which has extensive applications such as fluid analysis [1], chemical engineering [2], biological science [3] and others. Among these methods, digital holography (DH) is a non-intrusive and label-free method using a single image acquisition. Without using the lens, the interference pattern between the reference wave and the object wave through the particle volume is recorded by standard electronic sensors such as charge-coupled devices (CCDs). Afterwards, by back-propagating the holograms to a certain distance, the amplitude and phase information of the object optical field can be retrieved from the complex wave-front, which gives the 3D distribution of particles. Taking advantage from a simple and compact optical configuration, DH has been applied to characterize various particle fields such as biological cells [4, 5], spray droplets [6, 7], solid particles [8] and microplastics [9, 10] . Moreover, the trajectory of 3D particle path can be obtained by positioning particles detected in consecutive holograms, which forms the basis of holographic particle image velocimetry (HPIV) in diagnosing 3D particulate flows [11–14].

The traditional framework for extracting the 3D position of particles consists of two stages: image reconstruction and particle segmentation. The diffracted field is firstly reconstructed by numerically back-propagating the hologram at successive depth slices. Then, the particles are sized using advanced localization and segmentation algorithms. For example, the clustering-based particle detection (CBPD) method performs the K-means clustering to identify the regions of particles, particle edges and background after getting the reconstructed images at each depth slice from the hologram [15]. Tian *et al.* [7] project the minimum intensity location of the reconstructed planes along the axial direction at the pixel level, which is used as a focus metric to detect the focal plane of the bubbles. Then the edges of bubbles are further resolved by using morphological operators on the projection map. Although these previous studies have achieved significant progress, most of them exhibits poorer z-directional resolution than the transverse

xy-plane precision. The reason behind is the depth-of-focus problem, which is resulted from the small numerical aperture (NA) from the limited pixel size of sensors [16]. Moreover, especially in the inline (Gabor) holography set-up, the twin-image and self-interference-related spatial artefacts hinder the reconstruction of object in-focus with sharp edges and high spatial resolution. To tackle these problems, Molaei and Sheng [17] developed the correlation-based de-noising algorithm to clear the background noise and improve the hologram quality. Chen *et al.* [18] formed a denser point spread function by a spiral phase plate to enhance hologram reconstruction. Ren *et al.* [19, 20] combined the advanced data-driven method in auto-focusing and removing the undesired twin image.

However, the numerical reconstruction process from a captured hologram at consecutive depth slices is time-consuming and computationally demanding. Besides, the holograms processing amounted to localization problems, making it less necessary to realize image reconstruction with high quality. Therefore, some one-stage methods were proposed which skip the 3D reconstruction. The inverse methods attempt to find the optimal object 3D position that could produce the recorded interference pattern while satisfying some physical constraints [21]. For example, Verrier *et al.* [22] combined the inverse problem with pixel super-resolution to track colloidal particles in Brownian motion accurately. Mallery and Hong [23] proposed a regularized inverse method based on the fused lasso to introduce the smoothness and sparsity constraints into holographic volume reconstruction. However, the iterations in inverse methods still operate with considerable computation resources and stringent requirements, restricting their use in real-time applications and complex parametric models with dense particles.

Recently, deep learning techniques have been successfully applied in various tasks [24], which take advantage of a single feed-forward inference process. The combination of deep learning and DH has achieved encouraging results [25] in replacing numerical reconstruction and auto-focusing process [19, 26–28], improving holographic image resolution [29], classifying micro-plastics [9, 10] and depth estimation [30]. However, deep learning is still under-utilized for 3D holographic particle volume imaging, with only a few studies reported [31–35]. Shimobaba *et al.* [33] verified the feasibility of using modified U-Net to directly retrieve the particle 3D location from raw holograms without back-propagation. They represented the particle information by lateral position map and axial position map. The former denoted the particle's existence by binary values, while the latter indicated the depth as a 256-level grayscale image. However, the network suffers from a low extraction rate and nonphysical outputs. Shao *et al.* introduced the hologram formation knowledge into the network, such as maximum phase projections [31] and minimum intensity projection [32]. Although noticeable improvement has been reported in detection accuracy, the holograms have to be pre-processed beforehand to obtain the intensity and phase distribution of the optical field. Moreover, extra post-processing was required to transfer the network outputs (xy-centroid binary map and depth-encoded grayscale map) into 3D coordinates of particle, which is not intuitive and increases prediction time.

Therefore, in this work, we proposed a one-stage network (OSNet) for 3D particle volumetric reconstruction, which outputs the particle 3D location from raw 2D holograms in a single feed-forward process. Specifically, unlike other learning-based approaches using indirect maps, the 3D information of particle is represented by the centroid coordinate of the 2D bounding box in the transverse plane and normalized position along the axial direction, which is straightforward and avoids complex post-processing. The details of network structure and loss design will be introduced in Section 2. Then the detailed evaluation results will be presented in Section 3. Lastly, the conclusion and future work will be summarized in Section 4.

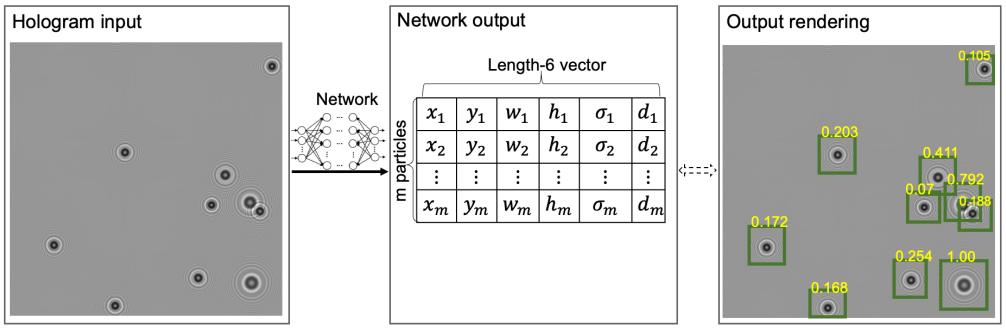

Fig. 1. The feed-forward inference process of OSNet. The predicted bounding boxes are rendered on the original hologram (rightmost figure) for better visualization. The number locating on the top of each bounding box is the normalized depth value for each particle.

2. Methodology

2.1. Particle characterization in one-stage detection framework

In the previous learning-based methods [31–33], the particle localization problem is considered as an image generation problem, since the 3D information of the particle volume is mapped to several two-dimensional images such as lateral and axial position maps. In that way, the additional post-processing operation is required to extract the 3D particle coordinates from the model output. Such operations are computationally demanding and time-consuming, which limit their use in real-time detection tasks. Besides, the target images are usually sparse because of the relatively small size of the particle centroid representation, which hinders the convergence of training process and results in making the false positive predictions i.e. the spurious particles [33].

Therefore, inspired by the YOLO framework [36] for object detection in computer vision, we annotate the 3D information of each particle by a 2D rectangle (bounding box). The central coordinate of the bounding box is the position of its enclosed particle in xy-plane and each box is assigned with a normalized depth value. Specifically, as shown from the schematic illustration of the network prediction process in Fig. 1, the 3D particle information is predicted from the OSNet by a set of fixed-length vectors $[x, y, w, h, \sigma, d]$. Each vector represents the bounding box geometry $[x, y, w, h]$, the objectiveness confidence σ and the normalized depth prediction d, respectively. It should be noted that, in order to make the network structure "particle-number agnostic" i.e. there is no assumption required about the particle number, a common strategy is to output the fixed number of bounding box proposals from the network [36]. After that, based on the predicted objectiveness score σ, the bounding box proposals with score higher than the threshold are selected as the final prediction. We suggest that, by migrating the 3D particle localization problem into bounding box regression problem, the detection accuracy will be improved with less non-physical output. Moreover, instead of considering the signal spread over the entire hologram, using the bounding boxes assists in more attention on local patterns surrounding the particles. This is more suitable for the holographic image processing because that the fringe patterns around the particles cover the main features for predicting the depth information. Following this one-stage detection framework, the OSNet directly retrieves the xy-centroid and depth location of particles from a single feed-forward process, which eschews the beforehand reconstruction and post-processing operations.

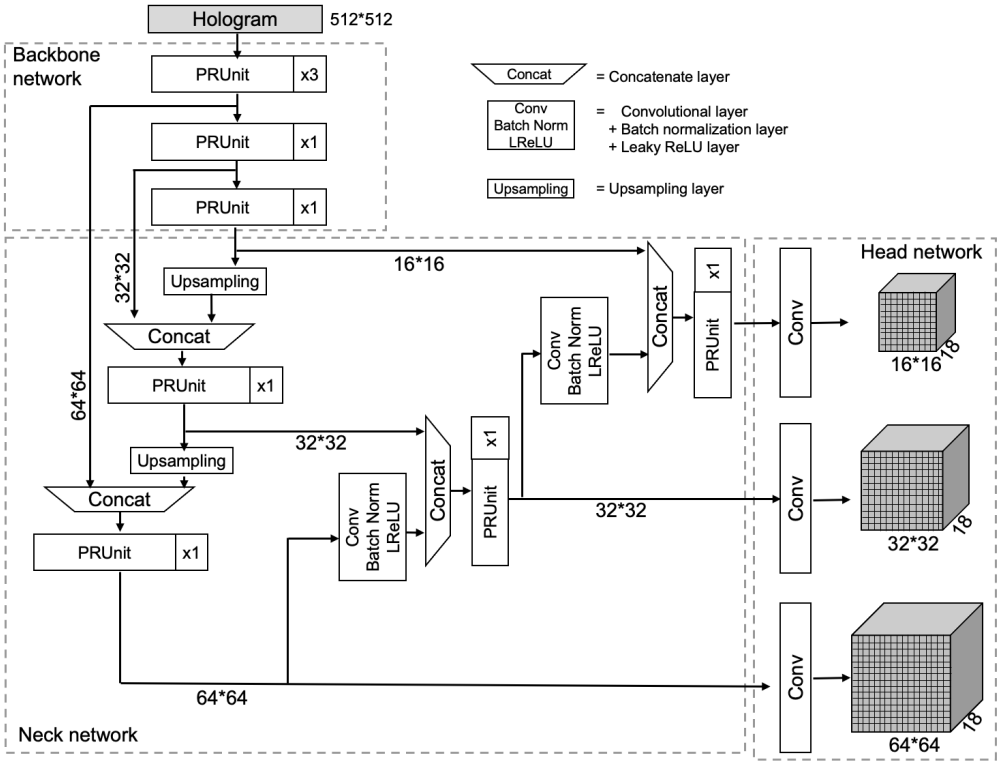

Fig. 2. OSNet architecture for particle detection which comprises three components: *backbone network*, *neck network* and *head network*. The number along the arrow connecting different blocks indicates the size of feature maps.

2.2. Network structure

The overall structure of OSNet is schematised in Fig. 2. The convolution operations in the convolutional neural network (CNN) provide affine invariant to recognize the contrast and spatially informative patterns within the hologram. Therefore, the input image is firstly passed through the backbone network, which comprises loads of convolutional layers to extract semantically strong features in a pyramid fashion. The representative power of the extracted features could be enriched by the number of stacked layers and the fusion between the consecutive feature maps. However, the network needs to consider the memory cost and the computational efficiency for 3D particle detection in real-time. Thus, the partial-residual unit (PRUnit) as shown in Fig. 3, is utilized as the major feature extraction blocks used in OSNet. It is adapted from the CSPNet [37], where the input feature is evenly split into two halves. One part is linked directly to the end of the stage after a dense block (convolutional layer + batch normalization layer + leaky ReLU layer). The other part is fed through consecutive dense blocks with a residual connection and concatenated with the first part. Such a partially processing strategy reduces the memory traffic and increases the gradient path, preserving feature reuse characteristics without duplicated gradient calculation. Besides, to avoid the performance degradation related to tiny particles, the feature map down-sampling is realized by another dense block instead of using the pooling layer.

In particle detection problems, the influence of the particle size needs to be considered. Since each cell in the feature map from different stages is responsible for a defined region of the input hologram, the combination of multiple resolution feature maps assists in detecting the particles with various size range. Thus, the neck network is followed, which aggregates feature maps

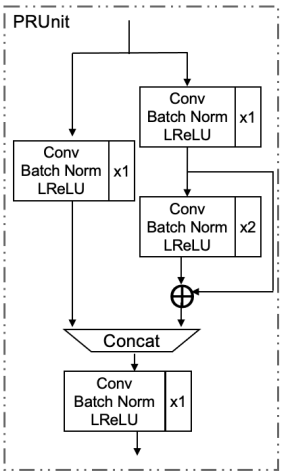

Fig. 3. The structure of the PRUnit in OSNet, which is the fundamental building block for feature extraction with enhanced learning capability and computational efficiency.

from different backbone stages to detect the components at various scales. The fusion method of neck network adopts a pyramid hierarchy with a top-down fashion [38], where the features from the current scale are concatenated with those from the previous scale. Considering the limited memory source, only the feature maps with size dimension ratio $\{\frac{1}{8}, \frac{1}{16}, \frac{1}{32}\}$ to the input image are taken. Moreover, the lower levels contain less semantic but spatially finer information, which may be discarded through the long propagation process. Therefore, an additional bottom-up path is utilised to shorten the information path of those information to the topmost features and enhance the final prediction of the large particles [39]. After passing through the neck work, the diverse information from different levels of the backbone network is incorporated, which strengthens the 3D information prediction of particles at various sizes.

Lastly, the bounding box proposals are predicted by the head network on feature map blocks, which are received from the neck network at each resolution level. Specifically, in order to get a more precise and efficient detection for objects of varying sizes, each feature map cell is responsible for predicting three bounding box proposals with different aspect ratios. Therefore, the output dimension of the head network is $m \times n \times (3 \times 6)$ for feature map block from neck network with size $m \times n$, which is achieved by pure convolution operations.

2.3. Loss design

As introduced in Section 2.1, the output bounding box prediction consists of a collection of fixed length vectors $[x, y, w, h, \sigma, d]$, which represents the bounding box geometry $[x, y, w, h]$, the objectiveness confidence σ and the normalized depth prediction d. Therefore, the overall loss for training the network is divided into three parts: the bounding box regression loss $\mathcal{L}_b$, the objectiveness loss $\mathcal{L}_{obj}$ and the depth regression loss $\mathcal{L}_d$. Each loss is designed and calculated independently. Then they are summed together as the final back-propagation loss for updating the parameters through the entire network as

$$\mathcal{L} = \mathcal{L}_b + \mathcal{L}_{obj} + \mathcal{L}_d. \tag{1}$$

Bounding box regression loss

In terms of bounding box regression, one of the trending evaluation metrics for predicted bounding box $B^p = [x^p, y^p, w^p, h^p]$ towards the target ground truth $B^t = [x^t, y^t, w^t, h^t]$ is the

Intersection over Union (IOU) score which is defined as

$$I_{(B^p,B^t)} = \frac{B^p \cap B^t}{B^p \cup B^t}, \tag{2}$$

where $B^p \cap B^t$ is the intersection area and $B^p \cup B^t$ is the union area of B^p and B^t respectively. However, IOU only considers the overlap extent while neglects the other geometry factors such as the central point distance and the aspect ratio. In order to address this drawback, the Complete Intersection over Union (CIOU) is proposed by Zheng *et al.* [40] as

$$C_{(B^p,B^t)} = I_{(B^p,B^t)} - \frac{\sqrt{(x^p - x^t)^2 + (y^p - y^t)^2}}{c^2} - \alpha v, \tag{3}$$

where c is the diagonal length of the smallest box enclosed by the union area, v is an estimation of the aspect ratio consistency and α is the positive trade-off parameter that are defined as

$$v = \frac{4}{\pi^2}\big(\arctan\frac{w^t}{h^t} - \arctan\frac{w^p}{h^p}\big)^2, \tag{4}$$

and

$$\alpha = \frac{v^2}{(1 - I_{(B^p,B^t)}) + v}. \tag{5}$$

As shown in Eq. (3), considering the three geometries (the overlap area, the centre point distance and the aspect ratio consistency) makes the CIOU more convincing to comprehensively evaluate the bounding box prediction performance. Therefore, the CIOU-based loss is adopted for the bounding box regression in our method, which is formulated as the mean CIOU value of each predicted bounding box with its target ground truth subtracted from 1

$$\mathcal{L}_b = \frac{1}{N}\sum_{i=1}^{N}(1 - C_{(B_i^p,B_i^t)}), \tag{6}$$

where N represents the number of bounding boxes that contain particles.

Objectiveness loss

The binary cross-entropy loss is used for objectiveness loss since it could be categorised as a binary classification problem

$$\mathcal{L}_{obj} = \frac{1}{N_p}\sum_{i=1}^{N_p}\left(\sigma_i^t \log(\sigma_i^p) + (1 - \sigma_i^t)\log(1 - \sigma_i^p)\right), \tag{7}$$

where N_p represents the total number of bounding box proposals, $\sigma^t \in \{0, 1\}$ is the ground truth label (1 for existence and 0 otherwise), and σ^p is the predicted objectiveness confidence ranging between 0 and 1.

Depth regression loss

The mean squared error (MSE) is utilized for depth estimation

$$\mathcal{L}_d = \frac{1}{N}\sum_{i=1}^{N}\|d_i^p - d_i^t\|_2^2, \tag{8}$$

where d^p is the predicted depth and d^t is the target ground truth for each bounding box with object. It should be mentioned that, for bounding box and depth regression loss, only the predicted bounding boxes with object occurrence (i.e. $\sigma^p > 0.5$) are considered since there is no information in background bounding boxes.

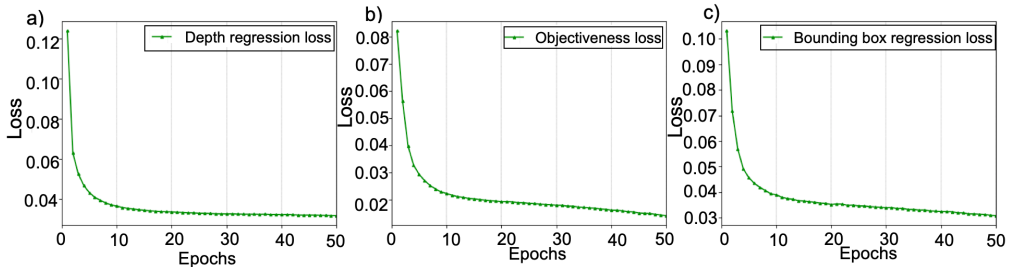

Fig. 4. Values of each loss component for the first 50 training epochs. (a) Depth regression loss. (b) Objectiveness loss. (c) Bounding box regression loss.

3. Experimental results and discussions

3.1. Dataset preparation

In order to verify the feasibility of our method, the synthetic holograms of the particles volume were generated following the inline DH propagation model introduced in [41]. The recorded hologram is formed by the interference between the object light $E_o(x, y)$ scattered from the particle field and the undisturbed reference light $E_r(x, y)$ passing through the volume, which is mathematically expressed as

$$I(x, y) = \mid E_o(x, y) + E_r(x, y) \mid^2 = \big(E_o(x, y) + E_r(x, y)\big)\big(E_o(x, y) + E_r(x, y)\big)^*, \tag{9}$$

where I is the intensity at each pixel, (x, y) is the pixel coordinate and $\{\cdot\}^*$ denotes the complex conjugate. Specifically, the particles were assumed to be pure amplitude disks with diameter ranging from 20 µm to 100 µm, which randomly distributed in a volume with distance 2 cm ~ 4 cm from the hologram plane. The image size was set to 512×512 with pixel resolution at 10 µm, which was in accordance with the common parameter of commercial image sensors. The wavelength of illuminating laser source was selected as $\lambda = 532$ nm. The synthetic holograms were normalized to the intensity range 0–255 in order to simulate the experimental holograms recorded by commercial CCD cameras.

3.2. Network training

The synthetic holograms were fed into the network with 3D information as labels for training. There were 10,000 holograms generated with particles concentration fixed at 1×10^{-3} particle/pixel (*ppp*). The depth label has been normalized to $[0, 1]$ in order to keep the network consistency and be able to generalize to different depth ranges. A typical split ratio of 7:2:1 was followed for training, validation and testing subsets, respectively. The network was implemented using Pytorch 1.8.1 and accelerated by NVIDIA Tesla V100 GPU. The Adam optimizer [42] was adopted with the initial learning rate of 0.01, which benefits in its ability in handling the sparse gradients and realizing efficient computation. We trained the network for 100 epochs with a batch size of 64 and cost 4 hours in total. The training losses during the first 50 epochs were plotted in Fig. 4. As illustrated, all the loss components decreased and converged gradually along the training process. This indicates that the proposed network is able to continuously update its parameters and retrieve 3D particle information directly from the raw holograms. Thus, the feasibility of our model was confirmed.

3.3. Performance evaluation on simulated data

The numerical evaluation for particle detection is commonly established on the extraction accuracy and the position prediction error. The extraction accuracy was assessed by precision p

Table 1. Evaluation results for particle extraction accuracy metrics and depth prediction error on test datasets with different particle densities.

Particle density (*ppp*)	Recall(%)	Precision(%)	Depth error(mm)
1×10^{-4}	98.34	99.52	0.17
2×10^{-4}	97.30	99.88	0.21
4×10^{-4}	97.30	98.20	0.31
6×10^{-4}	97.47	97.87	0.40
8×10^{-4}	96.92	98.29	0.49
1×10^{-3}	95.76	98.59	0.62

and recall r, which are defined as

$$p = \frac{N_t}{N_t + N_f}, \tag{10}$$

$$r = \frac{N_t}{N_t + N_m}, \tag{11}$$

where the N_t is the number of successfully paired particles, N_f is the number of false detection proposal and N_m is the number of unpaired ground truth (missed detection). The trained network reported 98.59 % recall and 95.76 % precision on 1000 test holograms with the same particle density as the training dataset (1×10^{-3} *ppp*). Regarding the position prediction, the average error was 0.62 mm, which accounted for 3 % of the distribution range along the axial direction (2 cm).

Performance on different particle concentrations

The robustness of our trained model with respect to various particle concentrations was then analysed on the test datasets with different particle densities ranging from 1×10^{-4} *ppp* to 1×10^{-3} *ppp*. Table 1 lists the accuracy metrics values and the mean depth prediction errors for each test dataset. Fig. 5 shows several selected scatter plots of the predictions and the ground truth. As demonstrated, the trained network showed the promising performance with above 95.76 % recall and 97.87 % precision at the varying densities. With the same particle density (1×10^{-4} *ppp*) and the depth range (2 cm) as [33], our method achieved 0.17 mm averaged positioning error in the axial axis, which is lower than 0.25 mm in their result. It shows that the attention on local patterns surrounding the particles has contributed to more accurate depth information retrieval from the hologram. For detection at a high-concentration 3D particle volume, Shao *et al.* [31] reported over 97 % detection accuracy at 1×10^{-3} *ppp*. In our method, a comparable result was obtained with 97.30 %. However, a higher precision 99.88 % was achieved by the OSNet, which means that there is a 0.12 % false prediction rate compared with 3.0 % in the literature. It indicates that the likelihood of making spurious particle predictions has been reduced by using the one-stage detection framework. One exemplary output under the denser condition with visualized bounding box proposals is shown in Fig. 6. As it suggests from the 3D rendering in Fig. 6(a), the predicted particles were generally in good alignment with the true distribution even under some conditions with dense spread (Fig. 6(d)). However, in the situation shown in Fig. 6(c), where the centroids of two particles were relatively close and there was a heavy disturbance between particle interference patterns, the particle centres may be hardly resolved. One possible solution we suggest is to introduce extra information such as multiple viewing angles offered in tomographic inline DH [13,43]. However, such supplementary

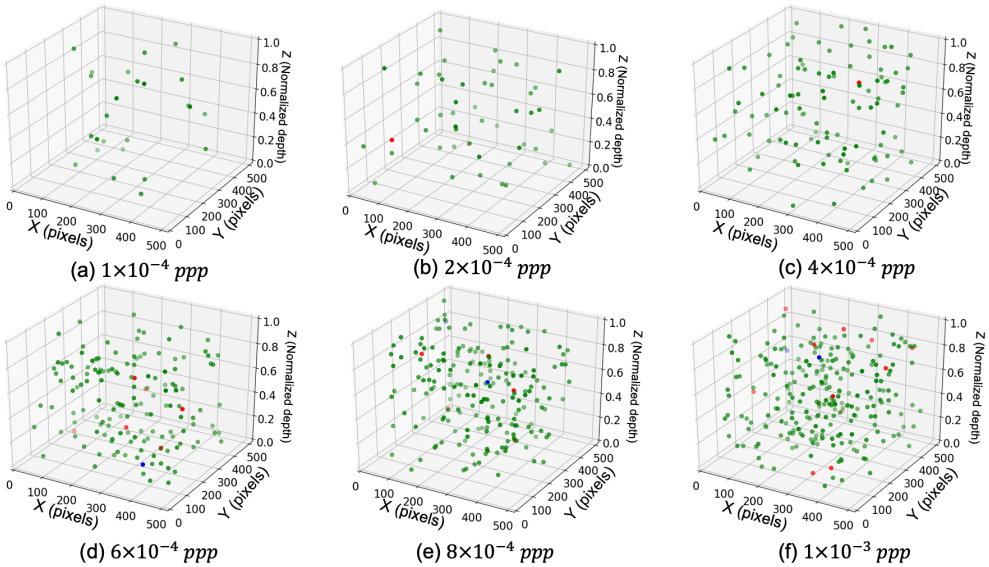

Fig. 5. Comparison of prediction results under different particle concentrations. The green dots are the correctly predicted particles. The red dots are the missed detection from the ground truth. The blue dots are the false detection.

information requires the additional burden of hardware such as mirrors or cameras, which harms the major advantages of inline DH in concise and compact set-ups.

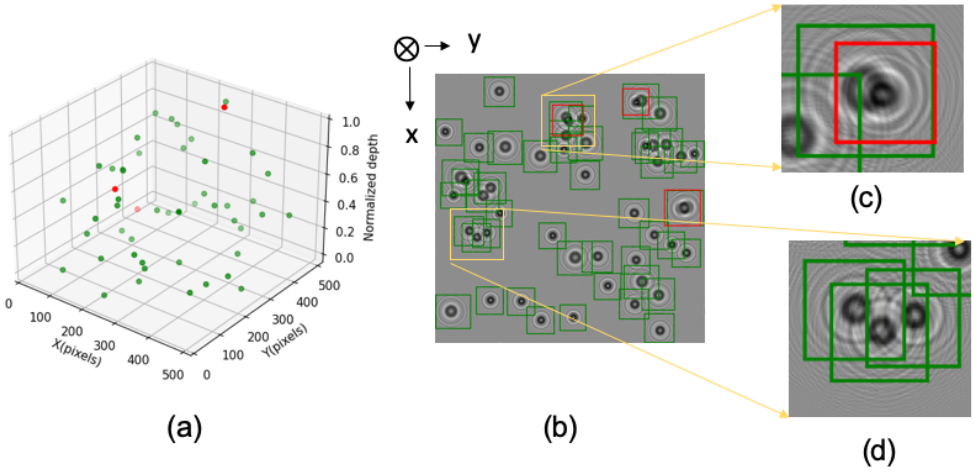

Fig. 6. Exemplary prediction result from the trained network with missed detection cases: (a) 3D visualization. The green dots are the correctly predicted particles and red dots are missed detection. (b) 2D visualization of bounding box predicted on holograms. The green boxes indicate the corrected proposed bounding box for true particles and red boxes are missed bounding box from the ground truth. (c) Missed detection. (d) Successful detection under dense distribution.

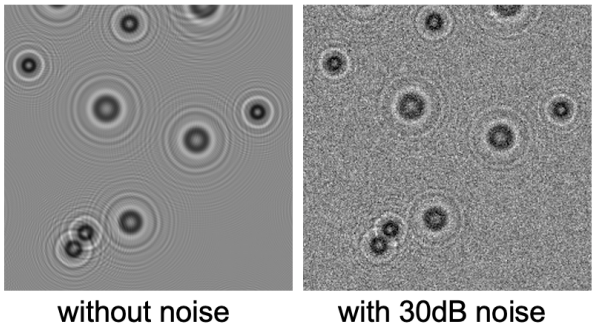

Fig. 7. Exemplary holograms without/with Gaussian noise.

Table 2. Evaluation results for particle extraction accuracy metrics and depth prediction error on test datasets with different noise level added.

Added noise level	Recall(%)	Precision(%)	Depth error(mm)
5 dB	98.42	99.28	0.16
10 dB	98.81	98.58	0.18
15 dB	98.62	98.73	0.25
20 dB	98.42	98.11	0.35
25 dB	98.62	99.28	0.45
30 dB	96.61	98.10	0.55

Performance on different noise levels

In order to verify the robustness of the developed model against the imaging noise, 10 groups of holograms were synthesized with varying Gaussian noise levels. The added noise was ranging from 5 dB to 30 dB with particle density 1×10^{-4} *ppp*. Fig.7 shows one example of hologram with 30 dB noise added, where the interference pattern is disturbed. The accuracy evaluation metrics and prediction error for each condition are listed in Table.2. Compared with the noise-free condition, the depth error showed slight increase from 0.16 mm with 5 dB added noise to 0.55 mm with 30 dB added noise of the entire distribution range (2 cm). It implies the inherent challenge in retrieving depth information from noise-disturbed interference fringes. However, the detection accuracy still maintained stable performance with above 96.61 % recall and 98.10 % precision over the studied range. Therefore, the robustness of the OSNet towards the imaging noise was verified with an acceptable performance decay in depth positioning accuracy.

Particle tracking performance

The accurate detection of the particle 3D location can be further utilized for particle tracking in space. To verify its feasibility, the helix movement of single particle in a volume was simulated in Fig. 8(a) and the hologram was recorded at one frame per movement step. Fig. 8(b) demonstrates the predicted result from our model, which shows general agreement with the real particle track in 3D space. For multiple particles tracking, the OSNet also preserved encouraging performance for particles which move randomly at the same time. Their true trajectories are plotted in Fig. 8(c) compared with their prediction in Fig. 8(d). The evidence from these evaluations intimates the extended application of OSNet in 3D particles tracking, which facilitates the analysis of the

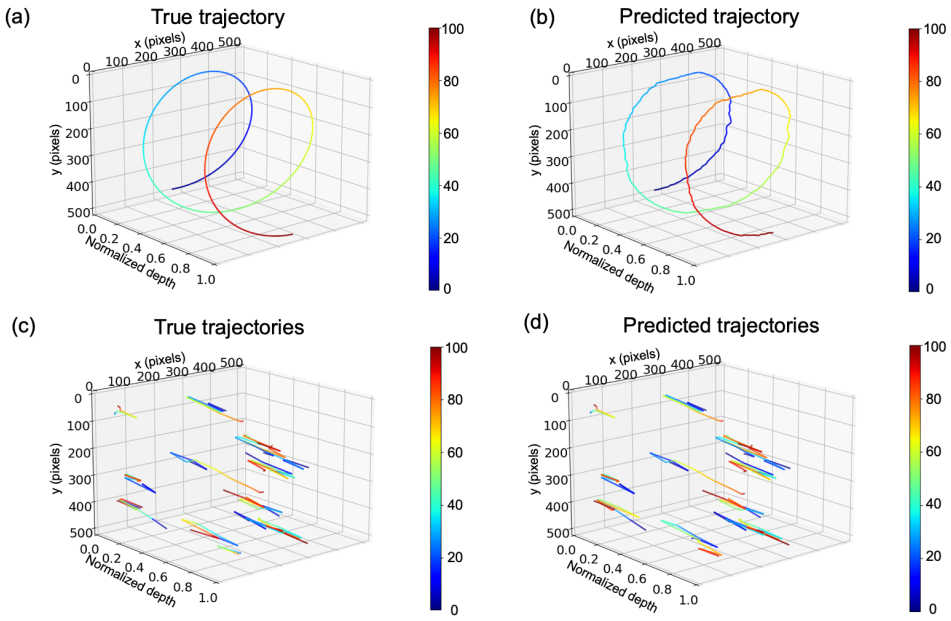

Fig. 8. 3D tracking result demonstration: (a)True trajectory for single particle which follows a helix curve. (b)Predicted single particle trajectory. (c)True trajectories for multiple particles with random movement. (d) Predicted multiple particle trajectories. The colorbar indicates the frame index of particle location, which is regarded as time step for the movement.

dynamic motions for micro-objects.

3.4. Performance evaluation on experimental data

The captured holograms in [34] were used to verify the feasibility of the OSNet on experimental data. In the experiment, the particles with a diameter of 50 µm were randomly seeded in water to generate the 3D particle field, which was placed at a distance ranging from 12 mm to 60 mm to the image sensor. The illuminating laser source used a wavelength of 660 nm and the holograms were recorded by the CCD with 3.45 µm pixel pitch. The pixel number was 1024×1024 which was cropped and scaled into 512×512 to match the network input size. The ground truth measurements were obtained by using regularzied inverse holographic volume reconstruction (RIHVR) [23], which is based on the traditional deconvolution with high-fidelity performance. For training the network, 400 randomly cropped holograms with the associated labels were used, and data augmentation was performed on the fly based on flip and rotation operations.

After training for 300 epochs, the prediction performance of the OSNet on experimental particle field is shown in Fig. 9. Since the particles were moving randomly within the volume, the captured holograms contained varying particle concentrations. The Fig. 9 (a)–(d) were the performance on sparse particle distribution with 22 particles. Most particle predictions were matched between the traditional deconvolution methods and our proposed model. However, with the similar reconstruction quality achieved, OSNet only took 21 ms while RIHVR took 17 s based on 300 iterations using the same computing hardware. For holograms with denser particle distribution ($\geq$ 50 particles) in Fig. 9 (e)–(h), the misalignment became more obvious especially along the axial direction. The major reason could be the resolution mismatch between the two methods. The resolution in traditional deconvolution method depends on the number of depth slices used in the model. Considering the computational burden and a reasonable inference

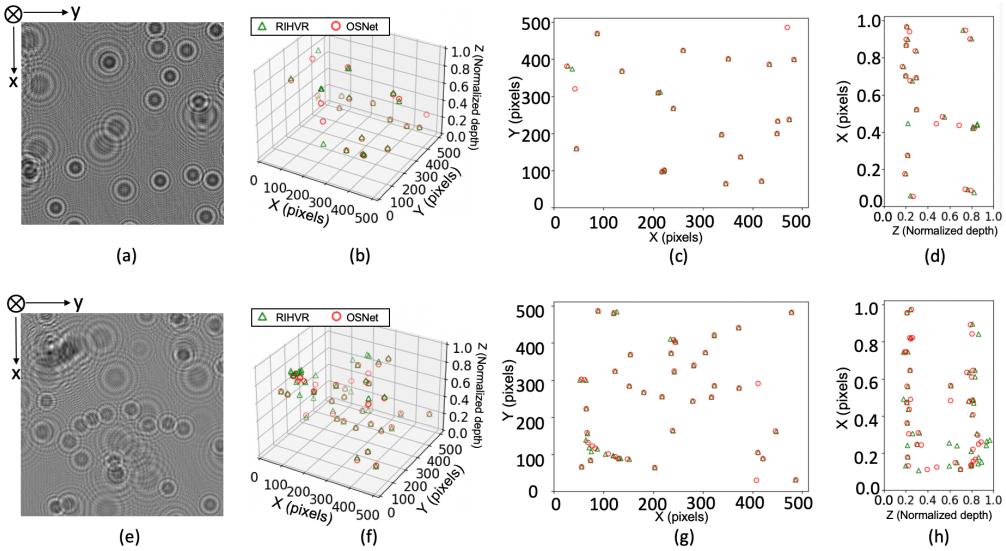

Fig. 9. The comparison of the experimental holograms reconstruction between the regularzied inverse holographic volume reconstruction (RIHVR) and OSNet. First row plots (a)–(d) are the comparison on a sparse particle field: (a) Original experimental recorded hologram. (b) 3D visualization of the reconstructed particle field from RIHVR (green triangles) and OSNet (red circles). (c) Top view comparison. (d) Side view comparison. Second row plots (e)–(h) are the comparison on a denser particle field: (e) Original experimental recorded hologram. (f) 3D visualization of the reconstructed particle field from RIHVR (green triangles) and the OSNet (red circles). (g) Top view comparison. (h) Side view comparison.

time, limited number of depth slices were used in the RIHVR that resulted in 1 mm resolution in this experiment. On the contrary, OSNet directly regressed the normalized depth value, which eschewed the resolution limitation associated with the number of depth intervals. Moreover, there were manually adjusted parameters such as the sparsity regularizer used in RIHVR for particle field reconstruction under different particle densities, while the learned weights used in OSNet were fixed once trained.

The holograms were recorded consecutively during a time interval in this experiment, thus the 3D particle tracking was performed by OSNet. For clarity, only a subset of the tracked particle trajectories was plotted in Fig. 10. Since the interference time of the proposed method was 21 ms, the highest frame rate that could be realized was 47 frame per second. Therefore, the feasibility of OSNet in providing the real-time particle tracking with an acceptable frame rate has been verified.

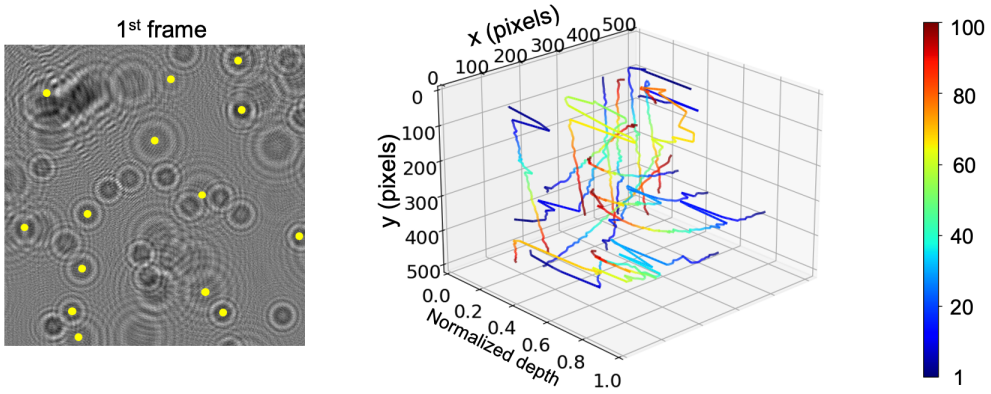

Fig. 10. Subset of tracked particles. The *left* figure shows the 1^{st} recorded frame of the particle field, where the yellow dots indicate the sampled particle centers. The *right* figure demonstrates the associated trajectories. The colorbar indicates the frame index of particle location, which is regarded as the time step for the movement.

4. Conclusion

In this study, the OSNet was proposed in order to replace the reconstruction process in volumetric particle reconstruction based on DH. By a single feed-forward process, the OSNet is able to extract the 3D centroids coordinates of particles directly from raw holograms without time-consuming pre- and/or post-processing operations. Compared with other learning-based methods [31–33], the competitive results have been obtained both in detection accuracy and depth positioning error. Furthermore, the robustness of the proposed model against the particle concentration and imaging noise was also verified. Moreover, the ultimate goal of accurate volumetric reconstruction in 3D particle tracking was provided by our model, where a high alignment between true trajectories and predicted tracks was demonstrated. In the end, the feasibility of OSNet on particle field reconstruction and tracking from experimental holograms has been validated. To our best knowledge, this is the first attempt to adapt the one-stage object detection framework into a 3D particle localization problem based on raw holograms with designed modifications. However, like most learning-based approaches, a group of labelled data is required as the ground truth for network training, which is labour-intensive and even not applicable in some cases. Thus, the ongoing work shall be devoted to introducing some physics-aware knowledge in the OSNet as an attempt to relieve the dependence on a large training datasets and challenge the performance limitation related to denser particle distribution.

Funding. The work is supported in part by the Research Grants Council of Hong Kong (GRF 17201818, 17200019, 17201620).

Acknowledgements. The authors would like to thank Dr. Ni Chen for access to the experimental data on the particle field.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. M. Raffel, C. E. Willert, F. Scarano, C. J. Kähler, S. T. Wereley, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide* (Springer, 2018).
2. V. F. Rad, A. Babaei-Ghazvini, R. Jamali, I. Shahabi-Ghahfarrokhi, and A.-R. Moradi, "Digital holographic microscopy for real-time investigation of 3D microstructural dynamics of starch-kefiran-based nanocomposite," Appl. Opt. **60**, 4706–4715 (2021).

3. I. Moon, F. Yi, and B. Rappaz, "Automated tracking of temporal displacements of a red blood cell obtained by time-lapse digital holographic microscopy," Appl. Opt. **55**, A86–A94 (2016).
4. P. Langehanenberg, B. Kemper, D. Dirksen, and G. Von Bally, "Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging," Appl. Opt. **47**, D176–D182 (2008).
5. P. Memmolo, M. Iannone, M. Ventre, P. A. Netti, A. Finizio, M. Paturzo, and P. Ferraro, "On the holographic 3D tracking of in vitro cells characterized by a highly-morphological change," Opt. Express **20**, 28485–28493 (2012).
6. Y. Yang and B.-s. Kang, "Digital particle holographic system for measurements of spray field characteristics," Opt. Lasers Eng. **49**, 1254–1263 (2011).
7. L. Tian, N. Loomis, J. A. Domínguez-Caballero, and G. Barbastathis, "Quantitative measurement of size and three-dimensional position of fast-moving bubbles in air-water mixture flows using digital holography," Appl. Opt. **49**, 1549–1554 (2010).
8. Y. Morita, T. Matsuo, S. Maeda, M. Oishi, and M. Oshima, "Three-dimensional displacement measurement of self-oscillating gel using digital holographic microscopy," Appl. Opt. **57**, 10541–10547 (2018).
9. Y. Zhu, C. H. Yeung, and E. Y. Lam, "Digital holographic imaging and classification of microplastics using deep transfer learning," Appl. Opt. **60**, A38–A47 (2021).
10. Y. Zhu, C. H. Yeung, and E. Y. Lam, "Microplastic pollution monitoring with holographic classification and deep learning," J. Phys. Photonics **3**, 024013 (2021).
11. J. O. Scherer and L. P. Bernal, "In-line holographic particle image velocimetry for turbulent flows," Appl. Opt. **36**, 9309–9318 (1997).
12. S. A. Wormald and J. Coupland, "Particle image identification and correlation analysis in microscopic holographic particle image velocimetry," Appl. Opt. **48**, 6400–6407 (2009).
13. J. Sheng, E. Malkiel, and J. Katz, "Single beam two-views holographic particle image velocimetry," Appl. Opt. **42**, 235–250 (2003).
14. Y. Wu, X. Wu, L. Yao, G. Gréhan, and K. Cen, "Direct measurement of particle size and 3D velocity of a gas–solid pipe flow with digital holographic particle tracking velocimetry," Appl. Opt. **54**, 2514–2523 (2015).
15. J. Huang, S. Li, Y. Zi, Y. Qian, W. Cai, M. Aldén, and Z. Li, "Clustering-based particle detection method for digital holography to detect the three-dimensional location and in-plane size of particles," Meas. Sci. Technol. **32**, 055205 (2021).
16. G. Pan and H. Meng, "Digital holography of particle fields: Reconstruction by use of complex amplitude," Appl. Opt. **42**, 827–833 (2003).
17. M. Molaei and J. Sheng, "Imaging bacterial 3D motion using digital in-line holographic microscopy and correlation-based de-noising algorithm," Opt. Express **22**, 32119–32137 (2014).
18. N. Chen, Z. Ren, H. Ou, and E. Y. Lam, "Resolution enhancement of optical scanning holography with a spiral modulated point spread function," Photonics Res. **4**, 1–6 (2016).
19. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica **5**, 337–344 (2018).
20. Z. Ren, Z. Xu, and E. Y. Lam, "End-to-end deep learning framework for digital holographic reconstruction," Adv. Photonics **1**, 016004 (2019).
21. J. Gire, L. Denis, C. Fournier, E. Thiébaut, F. Soulez, and C. Ducottet, "Digital holography of particles: Benefits of the 'inverse problem' approach," Meas. Sci. Technol. **19**, 074005 (2008).
22. N. Verrier, C. Fournier, and T. Fournel, "3D tracking the brownian motion of colloidal particles using digital holographic microscopy and joint reconstruction," Appl. Opt. **54**, 4996–5002 (2015).
23. K. Mallery and J. Hong, "Regularized inverse holographic volume reconstruction for 3D particle tracking," Opt. Express **27**, 18069–18084 (2019).
24. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).
25. T. Zeng, Y. Zhu, and E. Y. Lam, "Deep learning for digital holography: A review," Opt. Express (2021).
26. Z. Ren, Z. Xu, and E. Y. Lam, "Autofocusing in digital holography using deep learning," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV,* vol. 10499 (2018), p. 104991V.
27. Z. Ren, Z. Xu, and E. Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica **5**, 337–344 (2018).
28. T. Zeng and E. Y. Lam, "Model-based network architecture for image reconstruction in lensless imaging," in *Holography, Diffractive Optics, and Applications X,* vol. 11551 (2020), p. 115510B.
29. Z. Ren, H. K. H. So, and E. Y. Lam, "Fringe pattern improvement and super-resolution using deep learning in digital holography," IEEE Transactions on Ind. Informatics **15**, 6179–6186 (2019).
30. T. Pitkäaho, A. Manninen, and T. J. Naughton, "Focus prediction in digital holographic microscopy using deep convolutional neural networks," Appl. Opt. **58**, A202–A208 (2019).
31. S. Shao, K. Mallery, S. S. Kumar, and J. Hong, "Machine learning holography for 3D particle field imaging," Opt. Express **28**, 2987–2999 (2020).
32. S. Shao, K. Mallery, and J. Hong, "Machine learning holography for measuring 3D particle distribution," Chem. Eng. Sci. p. 115830 (2020).
33. T. Shimobaba, T. Takahashi, Y. Yamamoto, Y. Endo, A. Shiraki, T. Nishitsuji, N. Hoshikawa, T. Kakue, and T. Ito, "Digital holographic particle volume reconstruction using a deep neural network," Appl. Opt. **58**, 1900–1906 (2019).
34. N. Chen, C. Wang, and W. Heidrich, "Holographic 3D particle imaging with model-based deep network," IEEE

Transactions on Comput. Imaging **7**, 288–296 (2021).
35. N. Chen, C. Wang, and W. Heidrich, "Snapshot space–time holographic 3D particle tracking velocimetry," Laser & Photonics Rev. p. 2100008 (2021).
36. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* (2016), pp. 779–788.
37. C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,* (2020), pp. 390–391.
38. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* (2017), pp. 2117–2125.
39. S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* (2018), pp. 8759–8768.
40. Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," in *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 34 (2020), pp. 12993–13000.
41. J. W. Goodman, *Introduction to Fourier Optics* (W.H.Freeman & Company Ltd, San Francisco,CA,USA, 2017).
42. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations,* (2015).
43. J. Gao and J. Katz, "Self-calibrated microscopic dual-view tomographic holography for 3D flow measurements," Opt. Express **26**, 16708–16725 (2018).