PAPER • OPEN ACCESS

3D positioning and autofocus of the particle field based on the depth-from-defocus method and the deep networks

View the article online for updates and enhancements.

MACHINE LEARNING
Science and Technology

PAPER

3D positioning and autofocus of the particle field based on the depth-from-defocus method and the deep networks

Xiaolei Zhang[1,6], Zhao Dong[1,2,6] ⓘ, Huaying Wang[1,2], Xiaohui Sha[3], Wenjian Wang[4], Xinyu Su[1], Zhengsheng Hu[1] and Shaokai Yang[5,*]

1 School of Mathematics & Physics, Hebei University of Engineering, Handan, Hebei 056038, People's Republic of China
2 Hebei Computational Optical Imaging and Photoelectric Detection Technology Innovation Center, Hebei University of Engineering, Handan, Hebei 056038, People's Republic of China
3 Institute of Science and Technology Research, Hebei University of Engineering, Handan, Hebei 056038, People's Republic of China
4 School of Physics, Xidian University, Xi'an, Shaanxi 710000, People's Republic of China
5 Department of Physics, University of Alberta, Edmonton, Alberta T6GZE9, Canada
6 These authors contributed equally to this work.
* Author to whom any correspondence should be addressed.

E-mail: shaokai1@ualberta.ca

Abstract

Accurate three-dimensional positioning of particles is a critical task in microscopic particle research, with one of the main challenges being the measurement of particle depths. In this paper, we propose a method for detecting particle depths from their blurred images using the depth-from-defocus technique and a deep neural network-based object detection framework called you-only-look-once. Our method provides simultaneous lateral position information for the particles and has been tested and evaluated on various samples, including synthetic particles, polystyrene particles, blood cells, and plankton, even in a noise-filled environment. We achieved autofocus for target particles in different depths using generative adversarial networks, obtaining clear-focused images. Our algorithm can process a single multi-target image in 0.008 s, allowing real-time application. Our proposed method provides new opportunities for particle field research.

1. Introduction

Particle field positioning is a crucial task in various fields, such as biomedicine, materials science, and environmental engineering [1, 2]. Despite its importance, determining the three-dimensional (3D) position of each particle in a field remains a challenging task. While the lateral position can be obtained using centroid localization or object segmentation algorithms [3, 4], obtaining the longitudinal position is more challenging. Several methods have been proposed to address this problem, such as multiple particle-based imaging methods [5], precise calibration methods for vision measurement [6], and digital holography [7–9]. However, these methods have their limitations and are not always efficient or accurate.

The depth-from-defocus (DfD) method has also been used to determine the longitudinal 3D positions of the particle field. As early as 1987, Pentland used this method to investigate the depth of field [10]. Sometimes, there are defocus ambiguities in this method. Zhou *et al* used the DfD method with a single-lens dual-camera system to achieve the 3D positioning of moving particles. It solved the defocus ambiguity problem. However, the accuracy for the fuzzy particles was low [11]. Barnkob *et al* applied the defocus-particle-tracking (DPT) method to derive the depth coordinates of the particle images from different defocusing patterns. Two of the most common and widely used DPT approaches are based on the model functions (MFs) and cross-correlation (CC), respectively. MF-based methods work very well in low-concentration cases. CC-based methods have better robustness when the particle concentration is significant, and particle image overlap is vital [12]. Rossi and Barnkob applied this method to identify particles and estimate their 3D position. However, the iterative steps used here might reduce the processing

speed [13]. There needs to be further research on how to determine the 3D position of the particle field efficiently.

In recent years, deep learning (DL) has been applied for positioning. Franchini and Krevor used convolutional neural network (CNN) to the blur images obtained by an astigmatic system to improve the detection accuracy [14]. Nehme *et al* used CNN to extract the localization of 3D particles and dealt with 3D fluorescent microscopy for particles [15]. Dreisbach *et al* have recently utilized neural networks to enhance the rate of particle detection and reduce the occurrence of false positives, surpassing the capabilities of traditional detection algorithms [16]. Sachs *et al* have presented deterministic algorithms and deep neural networks that can recognize the size of up to four particle species simultaneously, with a particle diameter ranging from 1.14 μm to 5.03 μm [17]. Leroy *et al* used both the soft-assignment encoding and the DfD method to determine the intermediate depth for a single object from defocus blur images [18]. We have combined the DfD method with an efficient CNN called EfficientNet, which has demonstrated exceptional performance in image classification and object detection tasks [19]. However, a manual process is sometimes needed for lateral positioning. In this paper, our method overcomes the challenge of manual processes for lateral positioning by integrating you-only-look-once (YOLO), a state-of-the-art object detection framework that enables accurate and efficient lateral positioning of particles [20, 21]. Autofocus is another crucial aspect of particle field research [22–24], and we have incorporated generative adversarial networks (GANs) into our method to obtain clear and focused images of particles at different depths [25]. By combining these techniques, our proposed method enables accurate, efficient, and noise-robust 3D particle field positioning and autofocus. We believe that our method has significant potential to advance microscopic particle research and related fields.

2. Methods

2.1. Proposed method workflow and components

Our proposed approach for precise 3D particle field positioning, as illustrated in figure 1, utilizes the DfD method in combination with a powerful deep neural network-based object detection framework, YOLOv5. To prepare for training, we meticulously pre-processed the training images using labelImg software [26], a widely recognized image annotation tool in DL. Employing labelImg, we annotated the category name and location information for objects present in the training images. In this context, we considered depth as the category name and manually drew an anchor box for each sample with a known depth in the images. The lateral position was automatically derived from the box, and the corresponding information was converted into XML-format files, ready for YOLOv5 network training.
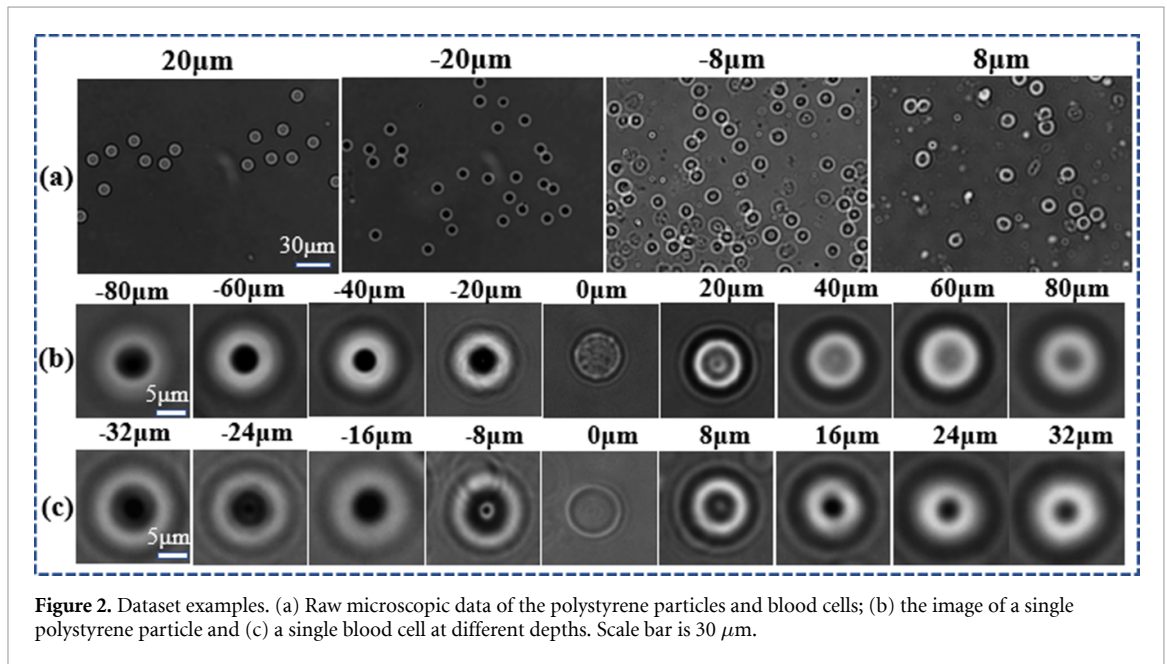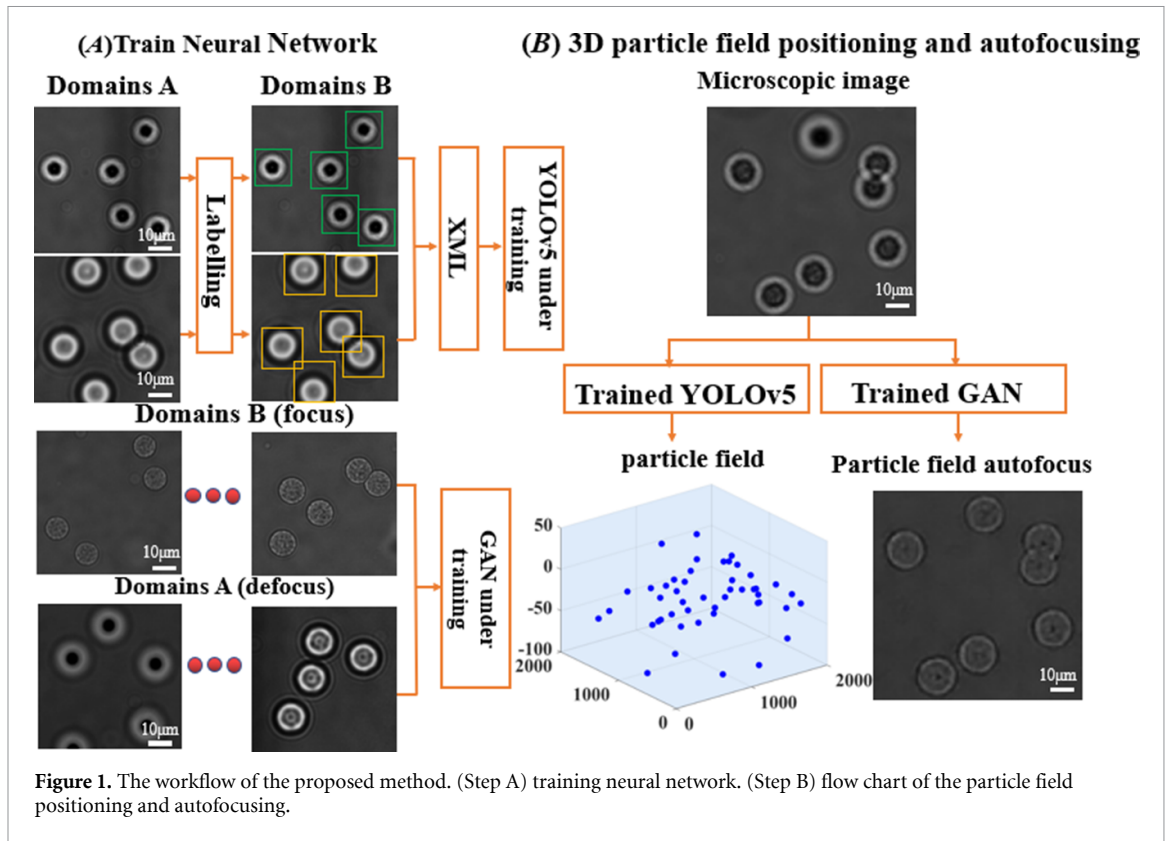
Simultaneously, we trained a GAN to obtain remarkably clear and focused images of the particles. We designated the defocus images as domain A and the focused images as domain B, utilizing both domains to effectively train the GAN. The expertly trained YOLOv5 network outputs the 3D positions of the particles, while the skillfully trained GAN generates focused images of the particle field. By seamlessly integrating YOLOv5 and the GAN, our proposed method achieves accurate, efficient, and noise-robust 3D particle field positioning and autofocus, representing a significant advancement in the field.

2.2. Experimental set-up

The experimental set-up for our proposed method involved obtaining micrographs of polystyrene particles as training input under a commercial microscope (XSP-37XF, Shanghai Opt. Inst. Fty., China). These particles were chosen for their mean diameter of approximately 10 μm and refractive index of 1.587. A 40$\times$ objective lens was used to capture their micrographs, and images for particles at different depths were obtained by carefully adjusting the longitudinal translation stage. The depths in focus were set to 0, while the depths for over-focus and under-focus conditions were greater and less than 0, respectively. The particles were placed under an Olympus objective lens oil during imaging, and some of the typical images obtained can be seen in figure 2. We have also tested our method on other particle fields, including plankton and red blood cells, and further details on the data collection process can be found in [19].

2.3. YOLO deep neural network and GANs

The YOLO network is a real-time object detection system that has shown great promise in accurately detecting object positions [27]. YOLOv5 network was the latest product of YOLO when we were performing our analysis, which has the advantages of fast detection and high accuracy [28]. It uses a one-stage neural network to complete detection object positioning directly. The model used here was modified from the YOLOv5s network. It has the smallest depth and width of the feature maps in the YOLO family [29], and the

Figure 1. The workflow of the proposed method. (Step A) training neural network. (Step B) flow chart of the particle field positioning and autofocusing.

Figure 2. Dataset examples. (a) Raw microscopic data of the polystyrene particles and blood cells; (b) the image of a single polystyrene particle and (c) a single blood cell at different depths. Scale bar is 30 μm.

detailed network structure can be found in [30]. Two types of GANs were utilized in our method for different conditions: Cycle-GAN and Pix2pix-GAN. Their structures and corresponding parameters can be found in [31].

The entire process was conducted on an Ubuntu 18.04 system. We trained our modified YOLOv5s network for 300 epochs using the Adam optimizer with a learning rate of 0.1 and a batch size of 2. The GANs were trained for 300 epochs with a batch size of 1 and a learning rate of 0.0001. The code for both networks was written in the Pytorch framework. We further processed the 3D images of the particles using MATLAB 2018a to enhance the clarity of the information presented. Our analysis was conducted on a desktop
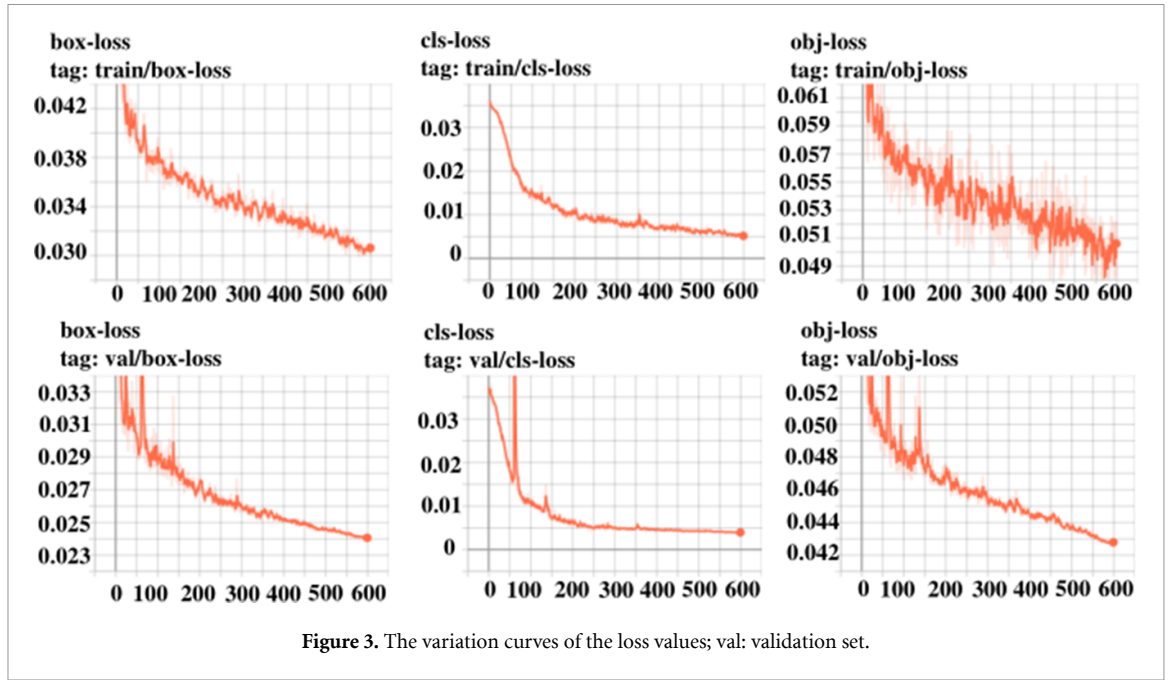
Figure 3. The variation curves of the loss values; val: validation set.

computer with an Intel Core processor i7-9700CPU, 3 GHz, GeForceGTX2060, and 8 GB video memory. The training data set and network code are available on GitHub https://github.com/xiaolei0828/particle-field.

2.4. Performance evaluation

The trained model can be evaluated using several parameters to assess its target detection ability. These parameters include precision (P), recall (R), average precision (AP), and mean average precision (mAP) [32]. The formulas for these metrics are defined as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

$$AP = \int_0^1 P(R)\,dR \tag{3}$$

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i. \tag{4}$$

Here, true positive (TP) represents the number of correctly detected target particles in the image, while false positive (FP) represents the number of false detections, and false negative (FN) represents the number of particles in the image that were not detected by the network. Precision is a metric that measures the ratio of correctly detected particles to all detected particles. It indicates how accurate the model is in detecting targets. Recall, on the other hand, measures the ratio of correctly detected particles to all particles in the sample. It shows how well the model can detect all instances of a target. AP is the average precision values obtained for different recall levels, and it is calculated by computing the area under the precision-recall curve. AP provides an indication of how well the model can detect targets at different levels of recall. mAP is the mean of the AP values for all classes, and it gives an overall measure of the model's performance.

2.5. Loss function of YOLOv5s

The YOLOv5s model employs a loss function consisting of three components: bounding box regression loss (box-loss), classification loss (cls-loss), and objectness loss (obj-loss) [33]. During the training process, monitoring the loss curves can indicate whether the network model is converging stably as the number of iterations increases. As shown in figure 3, the loss values decrease as the number of iterations increases when

Table 1. Comparisons of the training time and the parameters under different conditions.

Number of particles	Training		Parameters		
	Epochs	Time (h)	*P*	*R*	mAP
3361	200	0.946	0.748	0.944	0.905
	300	1.358	0.858	0.917	0.926
	600	2.758	0.864	0.935	0.905
5955	200	1.611	0.872	0.900	0.922
	300	**2.533**	**0.893**	**0.931**	**0.956**
	600	4.721	0.873	0.934	0.932
12 874	200	3.074	0.882	0.920	0.955
	300	4.599	0.888	0.921	0.955
	600	9.128	0.882	0.935	0.957

Note: Bold indicates the optimal performance values.

training and validating the model on polystyrene particles. The goal of this study is to estimate the depth class of the particles. The cls-loss curves become stabilized after 300 epochs, indicating that 300 epochs are sufficient for training the model to achieve stable convergence.

To determine the optimal number of iteration times and sample size for training, several tests were conducted using polystyrene particles as an example. The study compared the training time and evaluation indices of YOLOv5s using different amounts of data at different iteration times. As shown in table 1, the model achieved excellent performance after completing 300 epochs using a training set of 5955 samples. The *P, R*, and mAP were 89.3%, 93.1%, and 95.6%, respectively. Considering computational costs, the results suggest that 300 iterations and approximately 6000 particles are sufficient for effective training of the model.
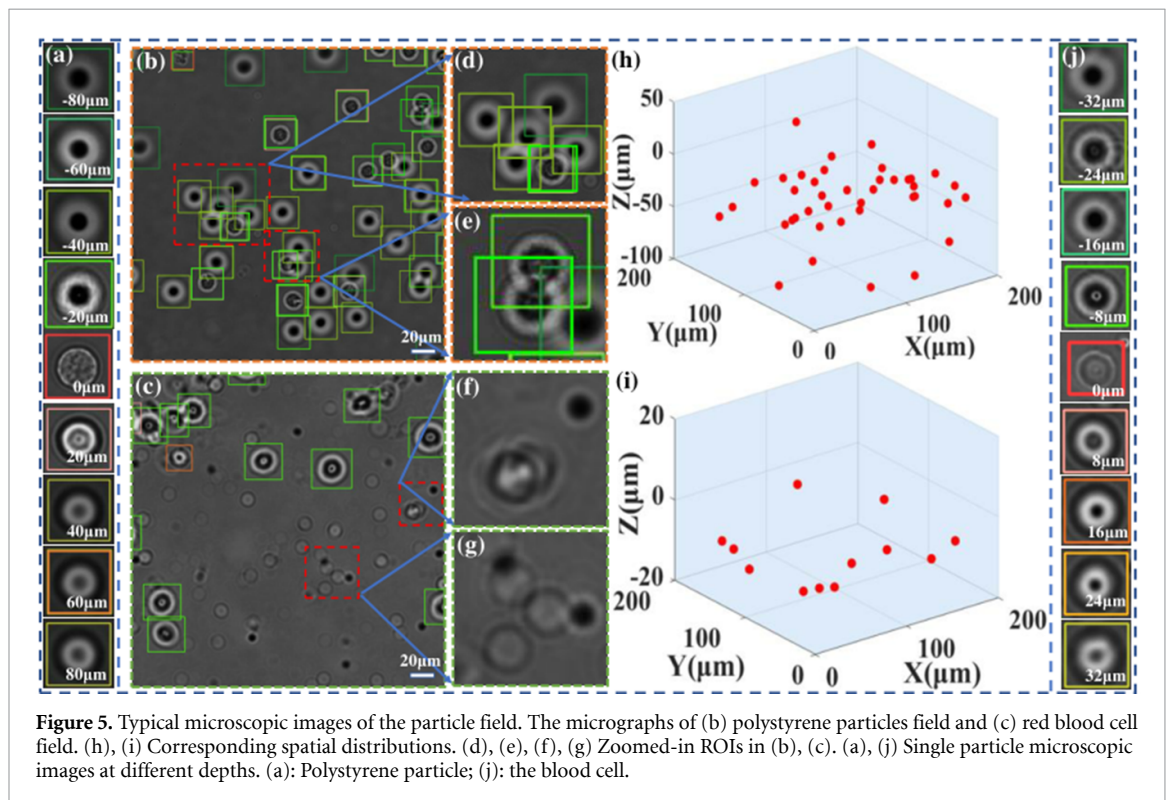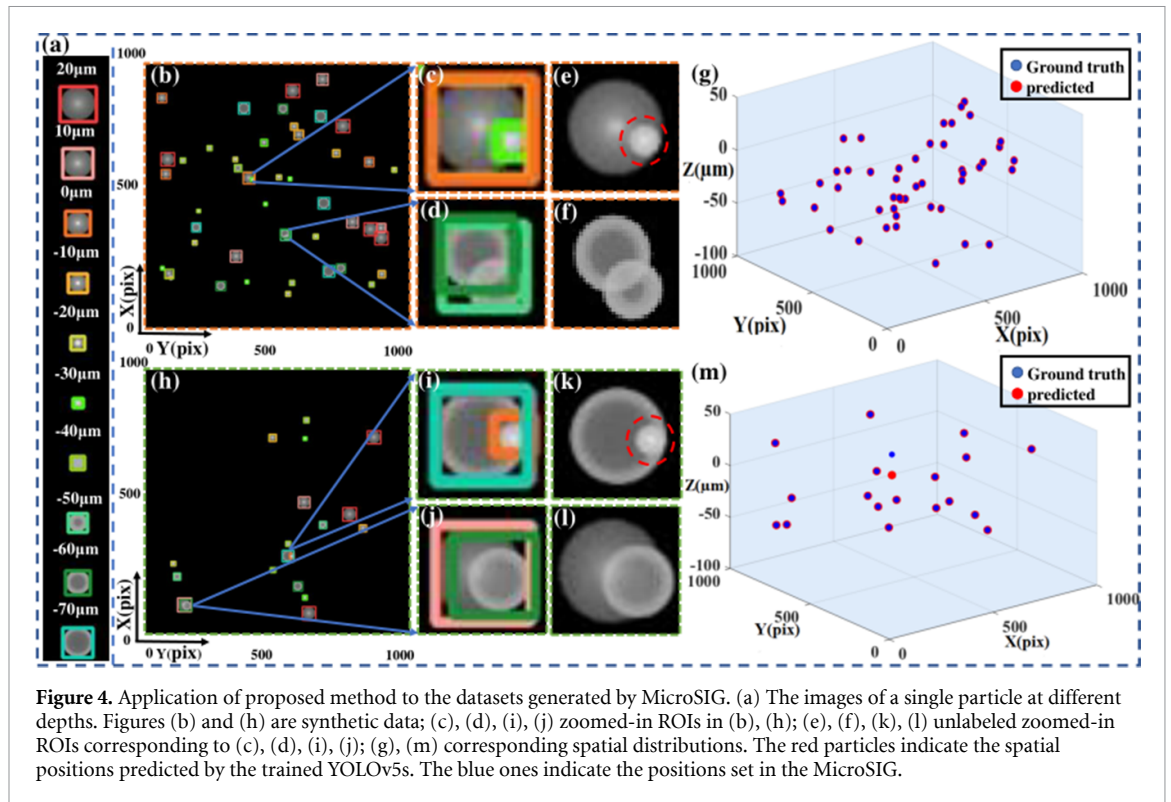
3. Results for the 3D positioning method

3.1. Validation on the synthetic dataset

The proposed method was validated using a synthetic dataset created by MicroSIG, a synthetic image generator based on a 3D ray-tracing approach proposed by Rossi [34]. Details about the synthetic dataset are available in the supplementary material. Figure 4 shows an example of the depth estimation on synthetic dataset. The output of the trained YOLOv5s network is shown in figures 4(b) and (h), where different colors indicate different depths. The 3D position information of each particle are stored in a file. Figure 4(a) shows the depth information with different color boxes. The enlarged regions of overlapping particles in figures 4(b) and (h) are shown in figures 4(c), (d), (i), and (j), respectively. Figures 4(g) and (m) show the 3D spatial distribution of particle fields corresponding to figures 4(b) and (h), respectively. The blue particles indicate the spatial positions of the particle field set by the MicroSIG, which are considered ground truths. The red particles indicate the positions predicted by the YOLOv5s network. The accuracy rate is approximately 99.9%, with most positions predicted correctly. However, the network needs better results for a very few particles. As shown in figures 4(e) and (k), the particles in the red dotted circle should have the same depth, but the network gives different depths. The correct depth of the particle is $-30~\mu$m. Mutual interference between overlapping particles may cause errors in figure 4(k). Nonetheless, this method can be successfully used in most situations, especially in sparse particle fields. The synthetic dataset has proven its feasibility, and the results suggest that the proposed method is effective in estimating particle depth.

3.2. Applications in the static field

The proposed method was applied to static particle fields using polystyrene particles and red blood cells as input examples. Five hundred-four images were used as the training input for the polystyrene particle field, and 5955 particles were extracted to train networks. The images were randomly divided into a training set and a verification set in an 8:2 ratio. To test the trained network, 1332 unlabeled particles with known depths were used, and the verification accuracy rate was approximately 99%. A similar process was performed for red blood cells using 308 images, with approximately 3447 cells for training and 230 for testing. The total accuracy rate was approximately 97.8%. Figure 5(b) shows a typical result for the polystyrene particles, while figure 5(c) shows the result for the blood cells. Each detected particle is bounded by a box, and its 3D information is output to a file. Different colors indicate different depths, as shown in figures 5(a) and (j). The corresponding 3D distribution can be seen in figures 5(h) and (i), respectively. Figures 5(d)–(g) show enlarged areas of overlapping particles and background noise. Even in the presence of overlapping particles

Figure 4. Application of proposed method to the datasets generated by MicroSIG. (a) The images of a single particle at different depths. Figures (b) and (h) are synthetic data; (c), (d), (i), (j) zoomed-in ROIs in (b), (h); (e), (f), (k), (l) unlabeled zoomed-in ROIs corresponding to (c), (d), (i), (j); (g), (m) corresponding spatial distributions. The red particles indicate the spatial positions predicted by the trained YOLOv5s. The blue ones indicate the positions set in the MicroSIG.

Figure 5. Typical microscopic images of the particle field. The micrographs of (b) polystyrene particles field and (c) red blood cell field. (h), (i) Corresponding spatial distributions. (d), (e), (f), (g) Zoomed-in ROIs in (b), (c). (a), (j) Single particle microscopic images at different depths. (a): Polystyrene particle; (j): the blood cell.

or cells and noisy environments, the method successfully detected and positioned them, demonstrating the robustness of the positioning method.

3.3. Applied to the dynamic field and planktons

The proposed method has also been applied to track the 3D motion of particles in a dynamic field. As an example, motion videos of polystyrene particles were captured and processed. The trained YOLOv5s network was able to automatically output the 3D positions of the target particles in real-time. A single frame with a
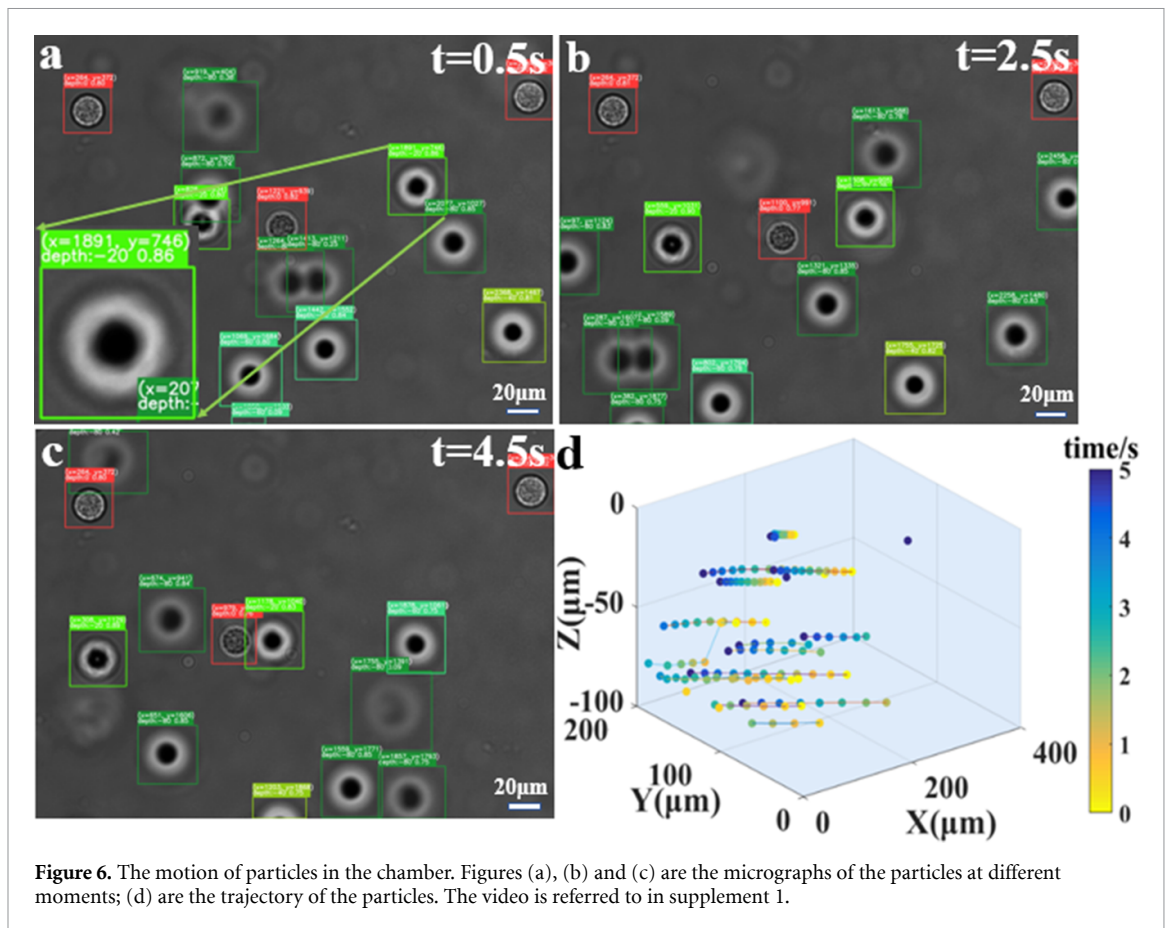
Figure 6. The motion of particles in the chamber. Figures (a), (b) and (c) are the micrographs of the particles at different moments; (d) are the trajectory of the particles. The video is referred to in supplement 1.

size of 1922 pixels * 2560 pixels was processed within about 0.008 s, as the video is usually recorded in 25 frames per second (fps), the 3D positioning of the particle fields can be completed in real-time. The results of the polystyrene particles in oil are shown in supplement 1 and in part in figures 6(a)–(c), while figure 6(d) shows the trajectories of the particles. The trajectories reveal that the oil mainly flows in a laminar manner and that the flow velocity may differ at different layers. The results demonstrate the great potential of this method for fluid investigation.

Furthermore, the proposed method was also applied to track the 3D motion of swimming planktons. Due to the difficulty of collecting moving underwater samples, the total number of samples was limited, and they were placed in a petri dish. Using the similar method introduced in section 2.2, microscopic images were captured using a 20×/0.40NA objective lens. Typical images of the planktons at different depths are shown in figure 7, while the results of two typical plankton types (named A and B) are presented in supplements 2 and 3, respectively. Part of these results can also be seen in figures 8 and 9. All samples were successfully detected and precisely positioned in the test process, and their 3D dynamic trajectory was obtained as shown in figures 8(d) and 9(d). The swimming planktons moved irregularly, and this method offered a new approach to the investigations of their behavior. It should be noted that in some cases, pollutants with similar color and morphology might be mistakenly considered as planktons, as shown in figure 9(c). This problem will be further discussed in the next section. Nonetheless, the results demonstrate the capability of this method to track the 3D motion of particles in a dynamic field, and it opens up new avenues for fluid and biological studies.

3.4. The influences of the noises, overlaps and the discrete depths

As previously mentioned, noise and overlaps from adjacent particles can introduce errors to the proposed method. To investigate the effects of these factors, polystyrene particles were used as an example, and the results can be seen in figure 10. The enlarged ROI 1 and ROI 2 in figures 10(b) and (c), respectively, show that noise does not affect the detection of target particles. The enlarged ROI 3 and ROI 4 in figures 10(d) and (e), respectively, demonstrate that most overlapped particles can be successfully detected and positioned, except for the particle in the center of the group composed of six-overlapped particles, which was affected by
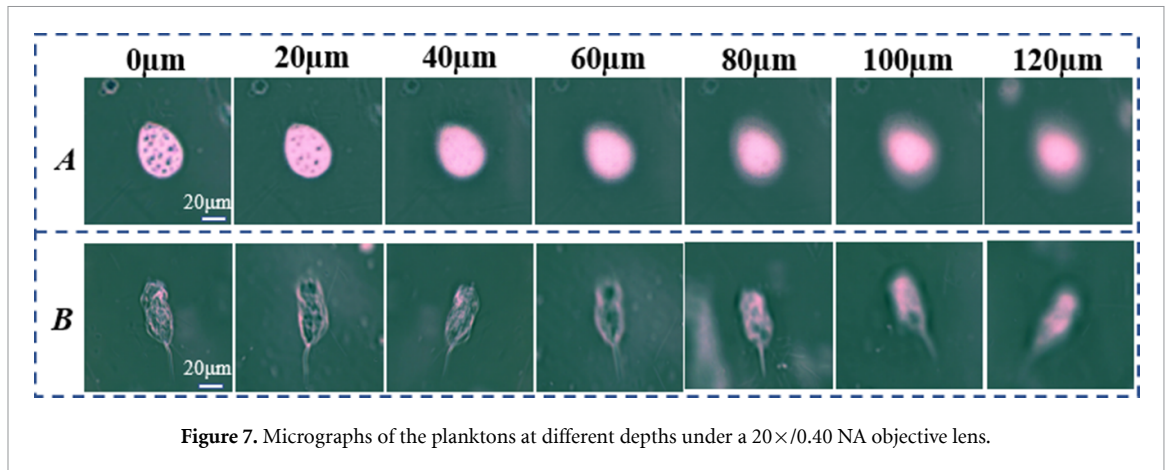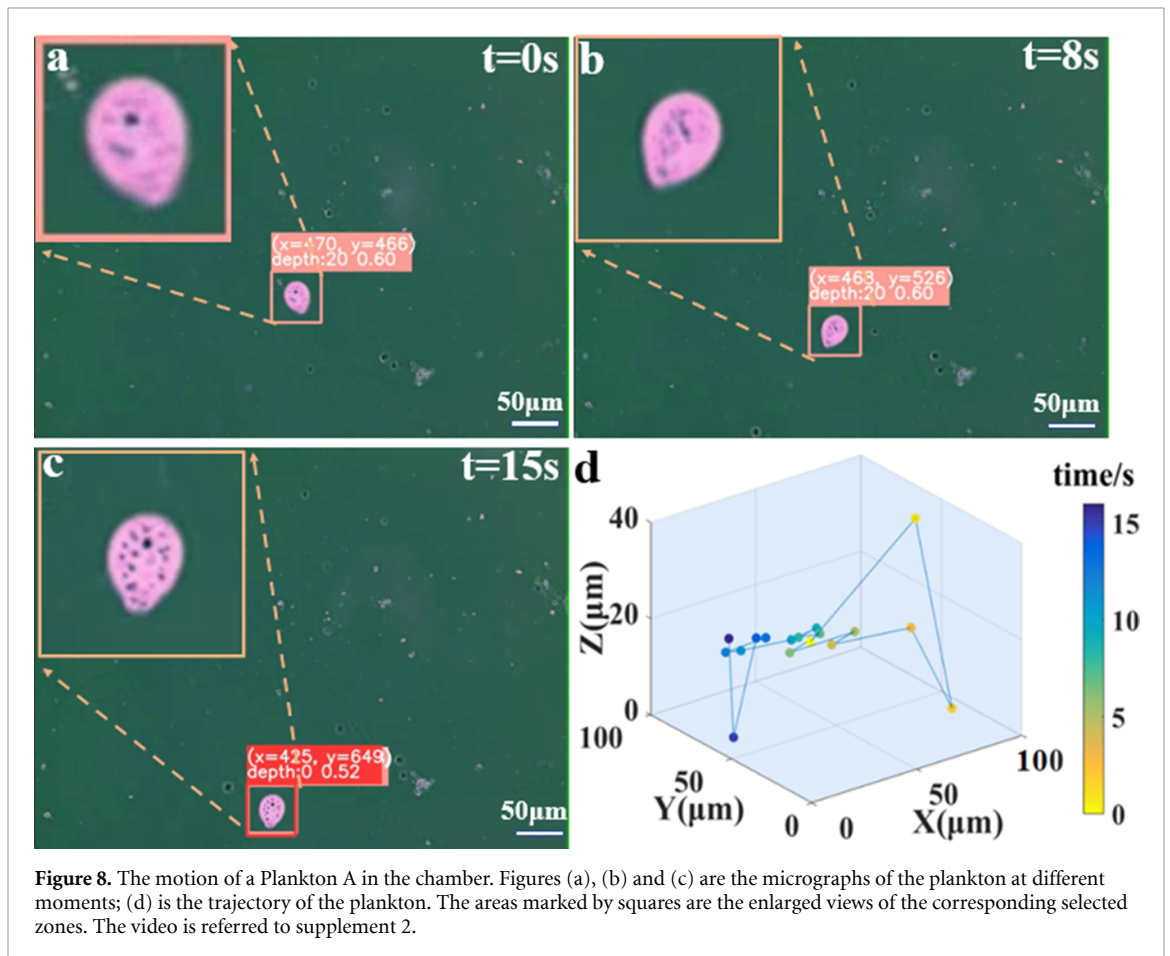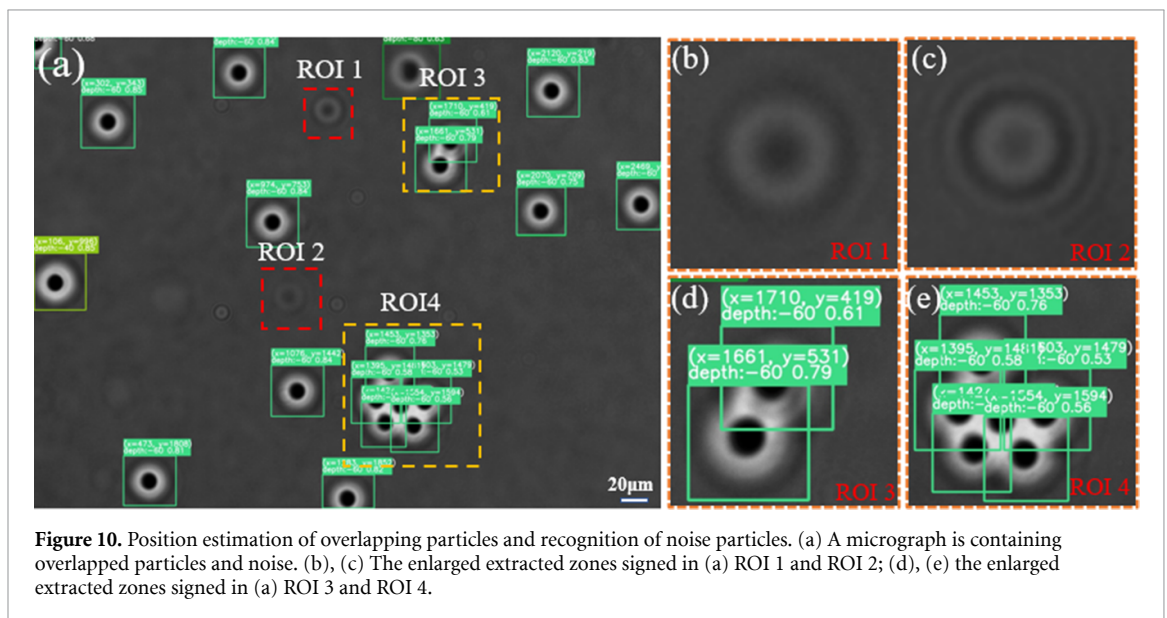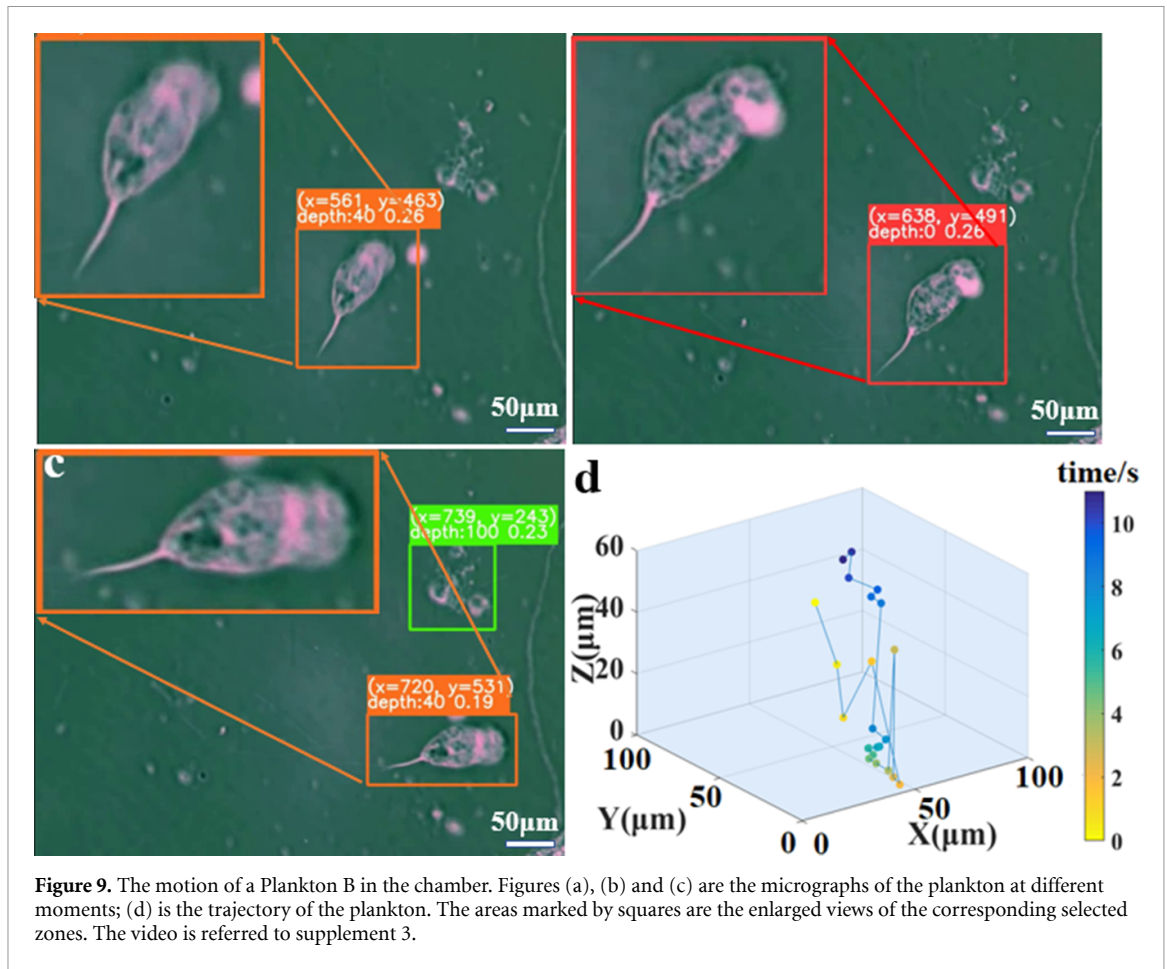
Figure 7. Micrographs of the planktons at different depths under a 20×/0.40 NA objective lens.

Figure 8. The motion of a Plankton A in the chamber. Figures (a), (b) and (c) are the micrographs of the plankton at different moments; (d) is the trajectory of the plankton. The areas marked by squares are the enlarged views of the corresponding selected zones. The video is referred to supplement 2.

the adjacent particles' morphology, leading to an error. However, as particles are usually dispersed before utilization, the proposed method is still suitable for most experimental conditions.

The depths determined by the proposed method are discrete, while the actual depth is a continuous variable. To obtain continuous depths, the network compares the input and images obtained at all set depths and calculates their corresponding probabilities, choosing the depth with the highest probability as the output. Continuous depths can then be calculated by taking the weighted average of all possible depths with their probabilities. While Leroy *et al* proposed a regression method to address the discretization of depth values [18], the adjacent particles' overlaps may still affect the probability distribution, and achieving accurate and consecutive depth estimation remains an ongoing research challenge.

Figure 9. The motion of a Plankton B in the chamber. Figures (a), (b) and (c) are the micrographs of the plankton at different moments; (d) is the trajectory of the plankton. The areas marked by squares are the enlarged views of the corresponding selected zones. The video is referred to supplement 3.

Figure 10. Position estimation of overlapping particles and recognition of noise particles. (a) A micrograph is containing overlapped particles and noise. (b), (c) The enlarged extracted zones signed in (a) ROI 1 and ROI 2; (d), (e) the enlarged extracted zones signed in (a) ROI 3 and ROI 4.

4. Autofocus

In situations where pollutants can lead to false detections in defocused images, autofocus is necessary. Therefore, the dataset prepared for 3D positioning can also be utilized for autofocus. However, DL-based autofocus methods require more data compared to 3D positioning methods and may require paired data. To address these issues, this section proposes the use of GANs, eliminating the need for additional data.
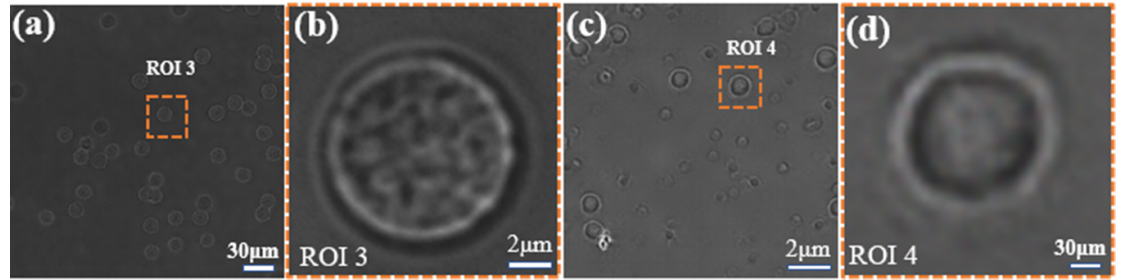
Figure 11. The autofocus of polystyrene particles and blood cells. (a) The focused polystyrene particle field and (b) the enlarged image of ROI 3. (c) The focused blood cell field and (d) The enlarged image of ROI 4.

4.1. GANs selection for autofocus

We have employed two types of GANs for autofocus, one that uses paired data and another that uses unpaired data. For the latter type, we used the widely-used Cycle-GAN [25]. It was used to complete autofocus for different fields, such as polystyrene particles and blood cells. The focus images of the target particles corresponding to figures 5(b) and (c) are shown in figures 11(a) and (c), respectively. Meanwhile, figures 11(b) and (d) show the focus images of the ROIs corresponding to figures 11(a) and (c), respectively. Compared to the images at different depths shown in figures 2(b) and (c), the particles and cells at different depths are now focused simultaneously. Cycle-GAN can learn the connections between two types of images from a large dataset [35], and since there is a large amount of data available for polystyrene particles and blood cells, Cycle-GAN works well in this scenario.

However, as the amount of plankton data is small, Cycle-GAN is unable to work effectively for this type of sample. Therefore, we opted for Pix2Pix-GAN to perform the autofocus. Pix2Pix-GAN requires paired data [36], where the input defocused image and the focused image (used as ground-truth) for training must be pixel-aligned. The network can learn the connections with relatively less data compared to Cycle-GAN.
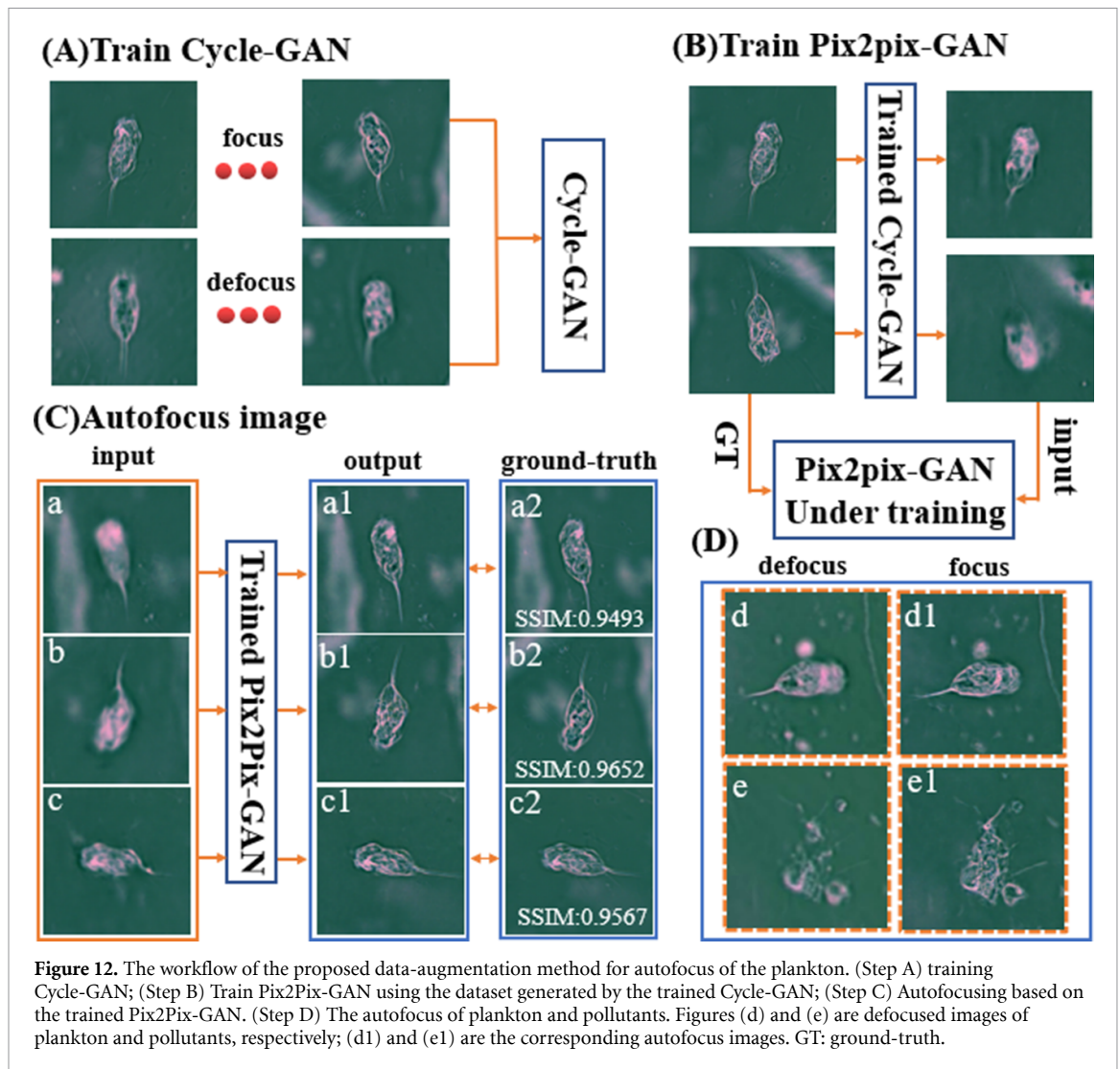
4.2. Data augmentation for autofocus of plankton

Because plankton samples are often in motion, obtaining paired data for autofocus is challenging. To address this issue, we extended our data augmentation method (as shown in [19] using Cycle-GAN. High-frequency information, such as details, is difficult to recover from a defocused image, but it is relatively easy to lose this information from a focused image. The workflow for this method can be seen in figure 12. We used Cycle-GAN, trained on a small amount of data, to generate defocused images at different depths from focused ones. These generated images and their corresponding focused counterparts were considered pixel-aligned and used to train the Pix2Pix-GAN. The trained Pix2Pix-GAN can then be used for autofocus. This method proved effective for plankton samples, where obtaining paired data is difficult due to their motion.

The structural similarity index measure (SSIM) [37] has been used here as an evaluation indicator to evaluate the accuracy of the images generated by the GANs. SSIM is defined as:

$$\text{SSIM}(a,b) = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{ab} + c_2)}{(\mu_a{}^2 + \mu_b{}^2 + c_1)(\sigma_a{}^2 + \sigma_b{}^2 + c_2)}. \tag{5}$$

Here, μ_a and μ_b are the mean values of images a and b, respectively. $\sigma_a{}^2$ and $\sigma_b{}^2$ are the variances, while σ_{ab} is the covariance of a and b; c_1 and c_2 are regularization parameters, respectively. A larger SSIM means a more substantial structural similarity between the two images. For example, if SSIM is 1, the two images are identical. The average SSIM values between the autofocused images and the corresponding ground truths are about 0.95. It indicates that this autofocus method is feasible. Three typical samples can be seen in figure 12. The SSIM values between the generated images (a1, b1, c1) and the corresponding experimental ones (a2, b2, c2) are about 0.9493, 0.9652, and 0.9567, respectively. As shown in figure 10(d), the pollutant was mistakenly recognized as a plankton. The plankton and pollutant were focused based on the proposed method, as shown in figures 12(d1) and (e1), respectively. They can be distinguished easily as the details can be observed clearly

Figure 12. The workflow of the proposed data-augmentation method for autofocus of the plankton. (Step A) training Cycle-GAN; (Step B) Train Pix2Pix-GAN using the dataset generated by the trained Cycle-GAN; (Step C) Autofocusing based on the trained Pix2Pix-GAN. (Step D) The autofocus of plankton and pollutants. Figures (d) and (e) are defocused images of plankton and pollutants, respectively; (d1) and (e1) are the corresponding autofocus images. GT: ground-truth.

in the focused images. This data augmentation method is suitable for the autofocus of the moving plankton. It offers a chance for further investigation of the plankton and the pollutants.

5. Conclusions

The proposed method for 3D positioning and autofocus of the particle field is a novel and effective approach that combines the use of the DfD and YOLO network. The depth of the particles can be determined by their blurred defocused images, and then the complete 3D position information can be obtained based on the trained YOLO model. The proposed method can process a single image containing multiple particles in about 0.008 s, making it suitable for real-time detection and 3D positioning. Furthermore, GANs were introduced to perform autofocus on particles at different depths simultaneously. The proposed positioning-autofocus method was validated on various samples and has demonstrated its robustness to the overlaps of adjacent particles and noise. With this method, the 3D positions of particles in the field can be determined in real-time, and their focused images can be generated. The results of this study suggest that the proposed method has great potential for application in the particle field, microbiology, environmental science, and fluid investigation, among other related areas. Additionally, the method's ability to detect and autofocus particles in real-time makes it highly practical and useful for researchers and scientists working in these fields. In conclusion, the proposed method offers a highly effective and efficient approach for 3D positioning and autofocus of particles, providing valuable insights for further research in the field of particle analysis and investigation.

Data availability statement

Acknowledgments

Conflict of interest

The authors declare no conflict of interes

ORCID iD

Zhao Dong ● https://orcid.org/0000-0002-5849-1979

References

[1] Wu Y C, Shiledar A, Li Y C, Wong J, Feng S, Chen X and Ozcan A 2017 Air quality monitoring using mobile microscopy and machine learning *Light-Sci. Appl.* **6** 17046

[2] Yoon J, Jo Y, Kim M H, Kim K, Lee S, Kang S J and Park Y 2017 Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning *Sci. Rep.* **7** 1–10

[3] Gao Z, Wu Y and Bao Y 2018 Image analysis for in-line measurement of multidimensional size, shape, and polymorphic transformation of l-glutamic acid using deep learning-based image segmentation and classification *Cryst. Growth Des.* **18** 4275–81

[4] Tsalicoglou C and Rösgen T 2022 Deep learning based instance segmentation of particle streaks and tufts *Meas. Sci. Technol.* **33** 114005

[5] Cierpka C and Kähler C J 2012 Particle imaging techniques for volumetric three-component (3D3C) velocity measurements in microfluidics *J. Visual-Japan* **15** 1–31

[6] Cui Y, Zhou F, Wang Y, Liu L and Gao H 2014 Precise calibration of binocular vision system used for vision measurement *Opt. Express* **22** 9134–49

[7] Dan H C, Bai G W, Zhu Z H, Liu X and Cao W 2022 An improved computation method for asphalt pavement texture depth based on multiocular vision 3D reconstruction technology *Constr. Build. Mater.* **321** 126427

[8] Chen N, Wang C and Heidrich W 2021 Holographic 3D particle imaging with model-based deep network *IEEE Trans. Comput. Imaging* **7** 288–96

[9] Lee S J, Yoon G Y and Go T 2019 Deep learning-based accurate and rapid tracking of 3D positional information of microparticles using digital holographic microscopy *Exp. Fluids* **60** 1–10

[10] Pentland A P 1987 A new sense for depth of field *IEEE Trans. Pattern Anal. Mach. Intell.* **4** 523–31

[11] Zhou W, Jin N, Jia M, Yang H and Cai X 2016 Three-dimensional positioning method for moving particles based on defocused imaging using single-lens dual-camera system *Chin Opt. Lett.* **14** 031201

[12] Barnkob R, Cierpka C, Chen M, Sachs S, Mäder P and Rossi M 2021 Defocus particle tracking: a comparison of methods based on model functions, cross-correlation, and neural networks *Meas. Sci. Technol.* **32** 094011

[13] Rossi M and Barnkob R 2020 A fast and robust algorithm for general defocusing particle tracking *Meas. Sci. Technol.* **32** 014001

[14] Franchini S and Krevor S 2020 Cut, overlap and locate: a deep learning approach for the 3D localization of particles in astigmatic optical setups *Exp. Fluids* **61** 1–15

[15] Nehme E, Freedman D, Gordon R, Ferdman B, Weiss L E, Alalouf O and Shechtman Y 2020 DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning *Nat. Methods* **17** 734–40

[16] Dreisbach M, Leister R, Probst M, Friederich P, Stroh A and Kriegseis J 2022 Particle detection by means of neural networks and synthetic training data refinement in defocusing particle tracking velocimetry *Meas. Sci. Technol.* **33** 124001

[17] Sachs S, Ratz M and Mäder P 2023 Particle detection and size recognition based on defocused particle images: a comparison of a deterministic algorithm and a deep neural network *Exp. Fluids* **64** 21

[18] Leroy R, Trouvé-Peloux P, Le Saux B, Buat B and Champagnat F 2022 Learning local depth regression from defocus blur by soft-assignment encoding *Appl. Opt.* **61** 8843–9

[19] Zhang X, Wang H, Wang W, Yang S, Wang J and Lei J 2022 Particle field positioning with a commercial microscope based on a developed CNN and the depth-from-defocus method *Opt. Lasers Eng.* **153** 106989

[20] Qi J, Liu X, Liu K, Xu F, Guo H and Tian X 2022 An improved YOLOv5 model based on visual attention mechanism: application to recognition of tomato virus disease *Comput. Electron. Agric.* **194** 106780

[21] Lei F, Tang F and Li S 2022 Underwater target detection algorithm based on improved YOLOv5 *J. Mar. Sci. Eng.* **10** 310

[22] Subbarao M and Tyan J K 1998 Selecting the optimal focus measure for autofocusing and depth-from-focus *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 864–70

[23] Akpinar U, Sahin E and Gotchev A 2019 Learning optimal phase-coded aperture for depth of field extension *2019 IEEE Int. Conf. on Image Processing(ICIP)* (*Taipei, Taiwan, 22–25 September*) (IEEE) pp 4315–9

[24] Wu Y, Rivenson Y, Zhang Y, Wei Z, Günaydin H, Lin X and Ozcan A 2018 Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery *Optica* **5** 704

[25] Zhu J, Park T, Isola P and Efros A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *2017 IEEE Int. Conf. on Computer Vision (ICCV)* (*Venice, Italy, 22–29, October*) (IEEE) pp 2242–51

[26] Liu C, Wu Y, Liu J and Sun Z 2021 Improved YOLOV3 network for insulator detection in aerial images with diverse background interference *Electronics* **10** 771

[27] Wang Lei, Yan Wei Qi, and 2021 Tree leaves detection based on deep learning *Geometry and Vision* vol 1386 ed M Nguyen, W Yan and H Ho (Cham: Springer) pp 26–38

[28] Yan B, Fan P, Lei X, Liu Z and Yang F 2021 A real-time apple targets detection method for picking robot based on improved YOLOv5 *Remote Sens.* **13** 1619

[29] Yao J, Qi J, Zhang J, Shao H, Yang J and Li X 2021 A real-time detection algorithm for kiwifruit defects based on YOLOv5 *Electronics* **10** 1711

[30] Wu D, Lv S, Jiang M and Song H 2020 Using channel pruning-based YOLOv4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments *Comput. Electron. Agr.* **178** 105742

[31] Wang W, Wang H, Yang S, Zhang X, Wang X, Wang J, Lei J, Zhang Z and Dong Z 2021 Resolution enhancement in microscopic imaging based on generative adversarial network with unpaired data *Opt. Commun.* **503** 127454

[32] Shinde S, Kothari A and Gupta V 2018 YOLO based human action recognition and localization *Proc. Comput. Sci.* **133** 831–8

[33] Malta A, Mendes M and Farinha T 2021 Augmented reality maintenance assistant using YOLOv5 *Appl. Sci.* **11** 4758

[34] Rossi M 2019 Synthetic image generator for defocusing and astigmatic PIV/PTV *Meas. Sci. Technol.* **31** 017003

[35] Tang C, Zhang W, Wang L, Cai A, Liang N, Li L and Yan B 2020 Generative adversarial network-based sinogram super-resolution for computed tomography imaging *Phys. Med. Biol.* **65** 235006

[36] Isola J, Zhu T, Zhou A and Efros A A 2017 Image-to-image translation with conditional adversarial networks *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (*Honolulu, HI, USA, 21–26 July*) (IEEE) pp 5967–76

[37] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12