On the preparation of training data for 3D tracking of position and orientation of microswimmers using defocusing and deep learning.

Mohammad Mehdizadeh Youshanlouei[1,*], Federica Miano[2], Massimiliano Rossi[1]

1: Department of Industrial Engineering, Alma Mater Studiorum, University of Bologna, Italy
2: Centre for Ocean Life, National Institute of Aquatic Resources, Technical University of Denmark, Denmark
*Corresponding author: mohammad.mehdizadeh2@unibo.it

Keywords: Deep learning, 3D Particle tracking, Defocusing particle tracking, Microfluidics, Microswimmers.

ABSTRACT

The use of image analysis algorithms based on deep convolutional neural network (CNNs) is opening new possibilities in experimental techniques for fluid dynamics. In the field of microscopic 3D particle tracking based on defocusing, it has been shown on synthetic images that CNNs can be used to determine not only the depth position but also the orientation of non-spherical microscopic particles or specimens. However, the translation of such approaches to experimental setups is challenging, especially due to the difficulty of obtaining reliable labeled images to train the neural networks. In this work, we propose a method to obtain labeled images of microswimmers from conventional microscopy images, exploiting the properties of the swimming motion at low Reynolds numbers, where inertial effects are negligible. This method allows to obtain ground truth values of orientation and depth position of a microorganism swimming with a regular pace that can be used as training data for CNNs. The proposed methodology was used to obtain labeled images from experimental videos of the micro-organism *Euplotes Vannus*. The labeled images were used to train a VGG16 network, providing an estimated uncertainty in the determination of the orientation angles ψ, θ, ϕ, and the depth position Z of 2.4%, 1.4%, 2.3%, and 3.5%, respectively.

1. Introduction

The recent advancement and wide availability of Deep Learning (DL) algorithms have significantly impacted many measurement methods based on image analysis. DL refers here to algorithms based on deep convolutional neural networks (CNN), which can be trained to perform a large variety of tasks, such as pattern recognition, image classifications, data prediction, etc. In most cases, the training procedure relies on a supervised learning approach, where a large number of labeled datasets are needed.

DL approaches are becoming increasingly popular in the field of defocus particle tracking (DTP), a category of 3D single-camera particle-tracking methods in which tracer particles in a volume are

observed with an objective with a small depth of field, and the out-of-plane particle coordinates are determined by the defocusing patterns of the respective particle images. DPT methods are mainly used in microfluidics experiments, where multi-camera approaches are not possible, and the tracer particles are observed through microscope objectives with large magnification and a relatively small depth of field (Cierpka & Kähler, 2012).

Conventional DPT methods rely on monodispersed, spherical, tracer particles and use a suitable calibration procedure to link the defocused particle-image patterns to the corresponding depth coordinates. For these applications, DL approaches have not shown significant improvements so far with respect to state-of-the-art methods (Barnkob et al., 2021). On the other hand, DL has the potential to deal with problems that are very difficult to solve with conventional approaches. One example is the measurement of polydispersed particles: Ratz et al. (2023) recently showed that CNNs can be used to determine the position of polydispersed particles with four different diameters.

Another interesting application is the measurement of the position and orientation of non-spherical microscopic objects, which can be used for instance to study the motion of microswimmers. Rossi (2021) has recently shown on synthetic images that it is theoretically possible to determine the orientation and position of spheroidal particles from defocusing images using CNNs and supervised learning. However, the main obstacle to the application of these methods to real experiments is the necessity of acquiring sufficiently large labeled datasets to train the neural network, i.e., images of the microswimmer with known orientation and depth position.

In this work, we propose a procedure to determine a large number of training images of a microswimmer that can be used to train a CNN to successfully identify the position and orientation of different specimens of the same organism observed with the same optical setup. The procedure is general and can be used on conventional optical microscopes and different types of organisms. The only requirement is to obtain images of the organism swimming freely with a constant periodic beat. The procedure is described in detail in the Material and Methods section and Results are shown on the heterotrophic ciliate *Euplotes vannus* (Rode et al., 2022).

2. Materials and methods

2.1. Methodology for the determination of ground truth data from 2D images

To describe the position and orientation of the microswimmer, we need the coordinates of the center of mass (x, y, z) and some Euler angles (ψ, θ, ϕ) to describe the body orientation. We chose the Tait-Bryan convention, where the angles correspond to yaw, pitch, and roll, respectively (Figure 1a). The determination of the in-plane position (x, y) is easy to obtain with conventional image analysis approaches. Python-based OpenCV has been implemented in this work and both KCF

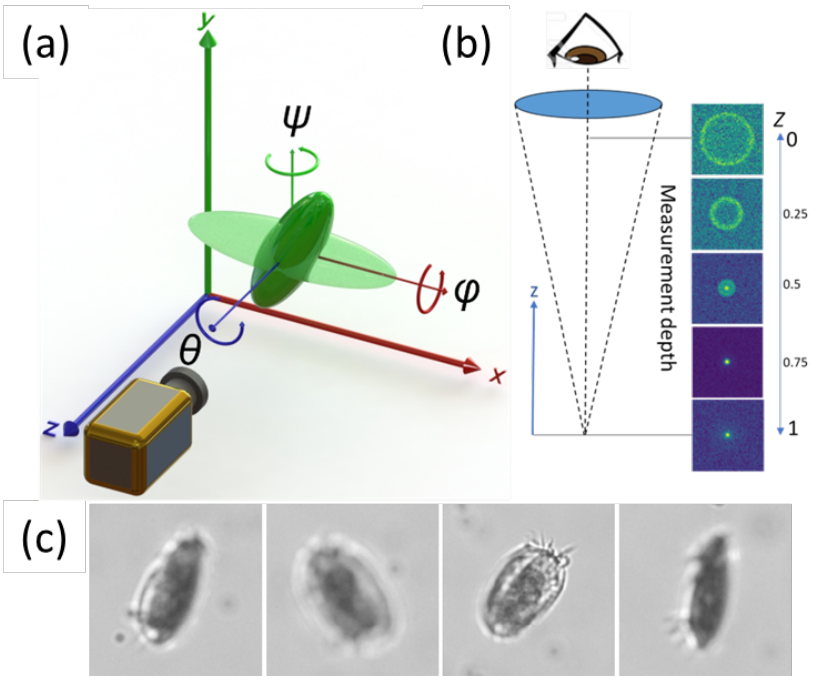

Figure 1. (a) Schematic of the conventions adopted to describe the position and rotation of a 3D object in this work. The Tait-Bryan convention was used for the Euler angles with yaw (ψ), pitch (θ), and roll (ϕ). z is the depth direction, perpendicular to the optical axis. (b) Principle of depth measurement using defocusing images for a single spherical particle. The depth coordinate is given in normalized unit $Z = z/h$, with h being the height of the measurement volume. (c) Exemplary images of the micro-organism *E. vannus* in different orientations and depth positions.

and CSRT trackers have been used to track the specimen in the frame. For frames belonging to different videos, both mentioned trackers were tested to achieve the most accurate result. DL is used to determine the depth position z and the Euler angles ψ, θ, ϕ from the defocused microscopic images of the microswimmer (Figure 1b-c).

To be able to train the CNN, one should collect images of the specimen at different depth positions and orientations. This is very challenging to achieve experimentally, however, in the case of microswimmers, it is possible to exploit a general characteristic of the motion at low Reynolds numbers (Purcell, 1977). When the microswimmer proceeds on a quiescent fluid at a constant pace (i.e., with a regular beat of its flagellum or cilia), its trajectory necessarily follows a helical path, and its body rotates around the axis of the helix. Under these conditions, it is possible to obtain the microswimmer 3D position and orientation from subsequent 2D image recordings (Rossi et al., 2017).

The first step is to obtain the specimen trajectory. We can write the equation of an arbitrary helical path using six free parameters: ρ, ω, and c to describe a general helix along the x-axis, and the Euler angles α, β, γ to properly orient the helix in space

$$\begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} + R(\alpha, \beta, \gamma) \begin{bmatrix} ct \\ \rho\cos(\omega t) \\ \rho\sin(\omega t) \end{bmatrix}, \tag{1}$$

where $R(\cdot,\cdot,\cdot)$ represents a rotation matrix as a function of three Euler angles. We can then use the x and y components of equation (1) to fit the experimental trajectory measured from the images ($x_m(t)$, $y_m(t)$) and thus obtain the trajectory ($x(t)$, $y(t)$, $z(t)$) of the specimen in the three-dimensional space (Figure 2a-b). The starting point can be chosen arbitrarily as $x_0 = x_m(0)$, $y_0 = y_m(0)$, $z_0 = 0$.

The next step is to obtain the orientation angles of the specimen as function of time. We follow here a trial-and-error approach. It should be noted that, under the current hypothesis of swimming motion, if the orientation angles are known for one time instant t_0, it is possible to obtain them for any time t just by applying an additional rotation around the helical axis u

$$R(\psi(t),\theta(t),\phi(t)) = R_u(\omega(t-t_0))R(\psi_0,\theta_0,\phi_0) \tag{2}$$

where $R_u(\cdot)$ represents a rotation matrix as a function of the rotation axis u and a given angle. We proceed by providing an initial guess of orientation angles ψ_0, θ_0, ϕ_0 that give a result in agreement with the experimental image at time t_0 (Figure 2c). We then derive the orientations for other frames from equation (2) and check the agreement with the experimental images (Figure 2b-c). We repeat the procedure until the optimal match is obtained. For this work, the agreement is evaluated from a qualitative comparison between a projected ellipsoid and the experimental images. In the future, more user-independent procedures based on some measurable metrics will be implemented.

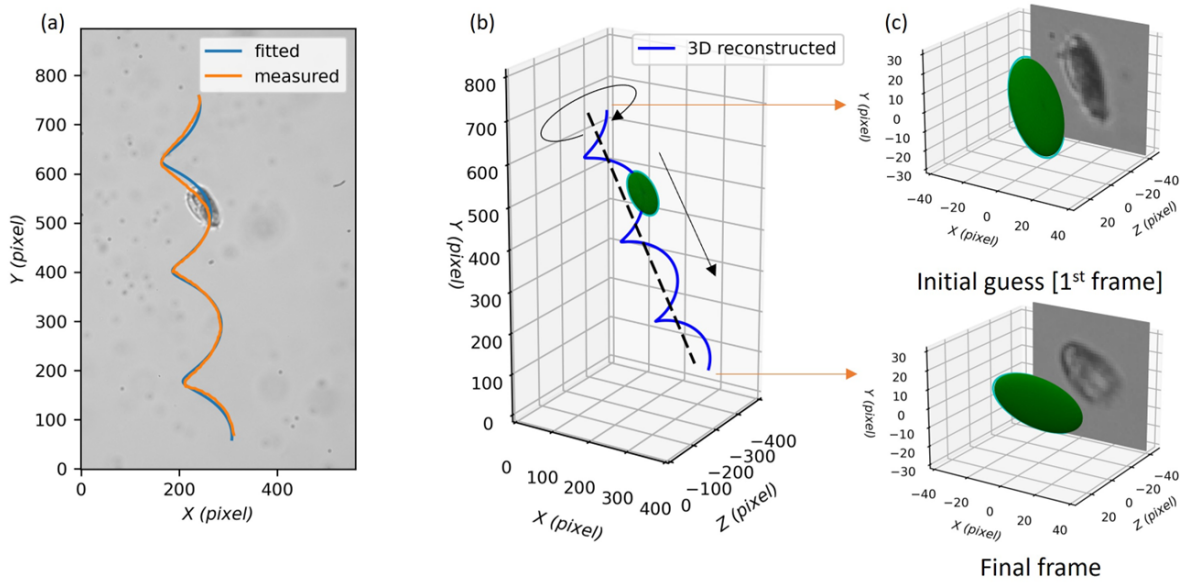

Figure 2. (a) Trajectory of the microswimmer measured with OpenCV and corresponding fitted trajectory using equation (1) in XY plane. (b) Reconstructed 3D trajectory of the microswimmer. The body rotates around the axis of the helix. (c) Trial-and-error procedure for the determination of the orientation. An initial guess is provided for a frame at $t = t_0$ and all the other orientations are obtained by applying a rotation around the helix axis.

2.2. Neural network architecture and training set

To achieve the optimal result, different CNN architectures with different modifications and settings were studied, namely ResNet18, ResNet50, ResNet101, and VGG16. Original versions of the mentioned architectures are designed for classification problems but the studied case is a regression problem so we had to modify the output layers of each model. Concretely, it means that each of the presented architectures will have a unique outlet box including a couple of Flatten, Dense, and Dropout layers. We used a custom input-image shape of $70 \times 70 \times 1$ pixels, (gray-scale images), therefore the use of a pre-trained model was not possible. Finally, the mean squared error for regression and Adam optimizer were applied. VGG16 with a learning rate of 0.0001 and batch size of 64 gave the best results in our case. The implemented version of VGG16 has one flatten layer followed by two Dense layers with size of 4096, Relu activation, and Dropout layers, and a final Dense layer with size of 4 and linear activation providing the output of Z, ψ, θ and ϕ. The depth position was given in normalized unit $Z = z/h$, where h is the height of the measurement volume.

The training data were obtained using selected frames from 5 videos of different specimens of *E. vannus* swimming at a regular pace, giving a total of approximately 500 frames. Since the frames belong to 5 different videos, overall we had 5 different background lights, noise, and contrast. In order to increase the number of training data, data augmentation has been applied. First, frames were randomly rotated and the corresponding rotation was applied to ψ. In the next step, the brightness and contrast of all images have been adjusted randomly in a specific range. As a result, a training set with almost 13000 images with an acceptable range of background settings was prepared.

2.3. Collection of images of *Euplotes Vannus*

The *E. vannus* culture (Culture Collection of Algae and Protozoa, SAMS Limited, Scottish Marine Institute, UK) was grown in artificial seawater at 18 °C and diluted 2–3 times per year in 65-mL flasks with artificial seawater and autoclaved rice grains to serve as a bacterial substrate. For the measurements, a group of different specimens was selected and placed in a custom-built, circular sample chamber and observed with an Olympus IX71 inverted microscope with a 10× objective. High-speed image recordings were obtained with a Phantom Miro LAB 320 at a frame rate of 150 fps.

3. Results

The VGG16 network with the modifications explained in section 2.2 was trained with the 13000 labeled images with orientation and Z position. All images were rescaled to be between 0 and 1. 5% of the images were separated randomly for testing and 5% for validation. 39,903,940 trainable parameters were trained for 37 epochs: Loss and validation loss are shown in Figure 3a. Overfitting

was observed during training which could be reduced by implementing the Dropout layers in the outlet box. Figure 3b shows the predicted versus true data for ψ, θ, ϕ, and Z. The Results indicate a mean average error of 2.4%, 1.4%, 2.3%, and 3.5% for ψ, θ, ϕ, and Z, respectively. As can be seen from the plot, there is a remarkable error at $\psi \approx 0$ and $\psi \approx 3.14$ which is due to the similarity of related images in these points. One critical aspect of the depth prediction is that the presented methodology provides values of z relative to an arbitrary position z_0, which might be different for different movies. In this work, we compare the defocusing patterns to align the depths of the different movies. In future studies, we are aiming to develop more robust strategies for improving this aspect. For instance, we are planning to use confined experimental setups where the specimen is constrained to swim within the measurement depth or to find other ways to obtain an independent measurement of the z values.

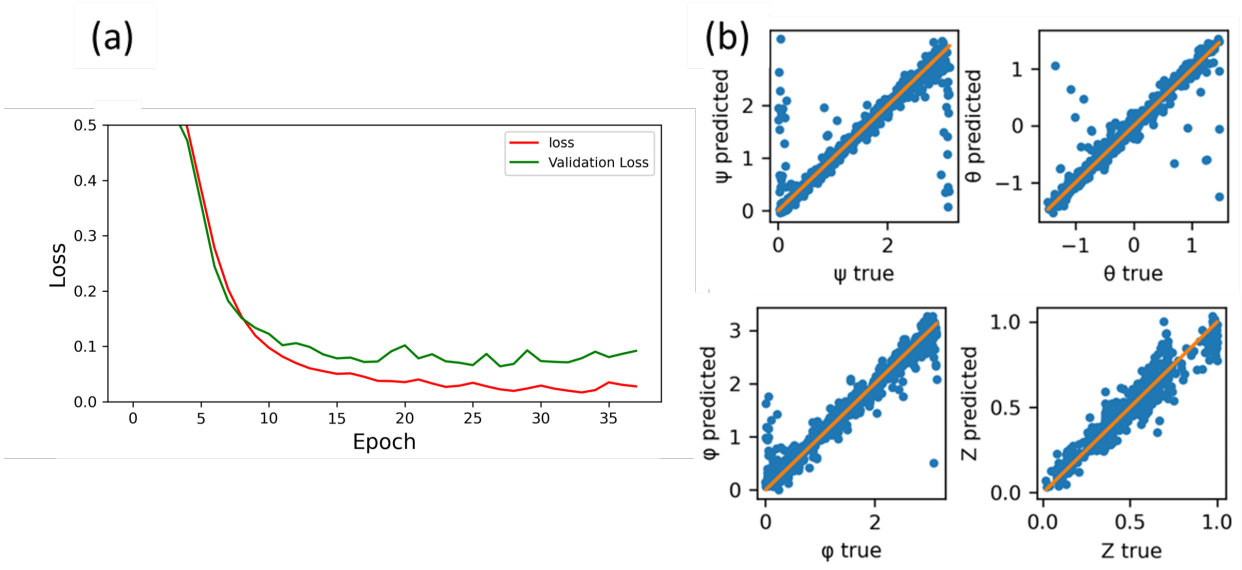

Figure 3. (a) The loss and validation loss for the training of the VGG16 network for 37 epochs. (b) True versus predicted values of ψ, θ, ϕ and Z for the 650 test images.

4. Conclusion

We provided a methodology to derive the 3D position (x, y, z) and orientation (ψ, θ, ϕ, corresponding to yaw, pitch, roll, respectively) of a microswimmer in motion with a regular pace on a quiescent fluid observed with a conventional microscope. The aim is to use these images to train CNNs able to measure the position and orientation of different specimens of the same organism swimming with an arbitrary motion. The CNN is used to measure the depth position z and the orientation angles ψ, θ, and ϕ. The in-plane positions x, y can be determined with conventional methods, in this work we used the Python-based package OpenCV. We apply the method on experimental images of the micro-organism *Euplotes vannus* that we used to train a VGG16 network.

After training, we obtain on test images an estimated uncertainty of 2.4%, 1.4%, 2.3%, and 3.5% for ψ, θ, ϕ, and Z, respectively. Follow-up works are planned to extend the size and quality of the training datasets and to apply the method to study the propulsion mechanisms of different micro-organisms.

Acknowledgements

MR acknowledges financial support from the VILLUM Foundation under the Grant No. 00036098. FM acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 955910. The authors thank Mads Rode for preliminary test images, and Anders Andersen and Thomas Kiørboe for helpful discussions.

References

Barnkob, R., Cierpka, C., Chen, M., Sachs, S., Mäder, P., & Rossi, M. (2021). Defocus particle tracking: a comparison of methods based on model functions, cross-correlation, and neural networks. *Measurement Science and Technology*, *32*(9), 094011.

Cierpka, C., & Kähler, C. J. (2012). Particle imaging techniques for volumetric three-component (3D3C) velocity measurements in microfluidics. *Journal of visualization*, *15*, 1–31.

Purcell, E. M. (1977). Life at low Reynolds number. *American journal of physics*, *45*(1), 3–11.

Ratz, M., Sachs, S., König, J., & Cierpka, C. (2023). A deep neural network architecture for reliable 3D position and size determination for Lagrangian particle tracking using a single camera. *Measurement Science and Technology*.

Rode, M., Kiørboe, T., & Andersen, A. (2022). Feeding flow and membranelle filtration in ciliates. *Physical Review Fluids*, *7*(2), 023102.

Rossi, M. (2021). On the determination of 3D position and orientation of spheroidal particles using defocusing and deep learning. In *14th International Symposium on Particle Image Velocimetry, Virtual Conference, August 1-4.*

Rossi, M., Cicconofri, G., Beran, A., Noselli, G., & DeSimone, A. (2017). Kinematics of flagellar swimming in Euglena gracilis: Helical trajectories and flagellar shapes. *Proceedings of the National Academy of Sciences*, *114*(50), 13085–13090.