Cut, overlap and locate: a deep learning approach for the 3D localization of particles in astigmatic optical setups

Simon Franchini¹ · Samuel Krevor¹

Received: 21 April 2019 / Revised: 17 March 2020 / Accepted: 30 April 2020 / Published online: 2 June 2020
© The Author(s) 2020

Abstract

Astigmatic optical systems encode the depth location of spherical objects in the defocus blur of their images. This allows the simultaneous imaging of 3D positions of a large number of such objects, which can act as tracer particles in the study of fluid flows. The challenge lies in decoding the depth information, as defocused particle images might be overlapping or have low maximum intensity values. Current methods are not able to simultaneously detect and locate overlapping and low-intensity particle images. In addition, their cost of computation increases with particle image density. We show how semi-synthetic images of defocused particle images with proximate center point positions can be employed to train an end-to-end trainable particle image detector. This allows for the detection of low-intensity and overlapping particle images in a single pass of an image through a neural network. We present a thorough evaluation of the uncertainty of the method for the application of particles in fluid flow measurements. We achieve a similar error in the depth predictions to previous algorithms for non-overlapping particle images. In the case of neighboring particle images, the location error increases with decreasing particle image center distances and peaks when particle image centers share the same location. When dealing with actual measurement images, the location error increases by approximately a factor of two when particle images share the same center point locations. The trained model detects low-intensity particle images close to the visibility limit and covers 91.4% of the depth range of a human annotator. For the employed experimental arrangement, this increased the depth range along which particle images can be detected by 67% over a previously employed thresholding detection method (Franchini et al. in *Adv Water Resour* 124:1–8, 2019).

✉ Simon Franchini
s.franchini15@imperial.ac.uk

¹ Imperial College London, London SW7 2AZ, UK

Graphic abstract

Generation of semi-synthetic training set with highly-overlapping particle images

Training of Deep Neural network in the detection and 3D localization of tracer particles

1 Introduction

Defocus blur is a widely researched image feature due to its property to encode depth information about objects in images, in this study particles entrained in a flow field. One application of this principle is the computation of depth maps from images acquired under different optical camera parameters (Guo et al. 2017; Subbarao and Surya 1994; Watanabe and Nayar 1998). In this way, cameras can act as depth sensors for arbitrary objects when lens, aperture, distance between lens and sensor or camera position are altered during image acquisition.

In the case of spherical objects or objects small enough to be approximated by a point, the depth location can be determined without the manipulation of the optics of the camera. Instead, the defocused image of the object can directly be used to infer the depth position of the object. A common application of this measurement principle is the field of microfluidics. 3D particle tracking velocimetry is a technique in which fluids are doped with spherical tracer particles. The tracking of the movement of the passive tracer particles reveals the motion of the fluid (Raffel

et al. 2018). This method is not limited to tracer particles. Actively moving objects such as bacteria might just as well be the object of investigation (Wu et al. 2005). Defocused images allow for the localization of such objects in 3D and therefore for the investigation of their motion in 3D.

Fig. 1 Experimental setup composed of the experimental scene which is magnified by a microscope objective and imaged by the camera sensor. The cylindrical lens creates a unique relation between the location of a spherical particle along the optical axis and its defocused image

To enhance the difference in the defocused images of particles along the optical axis, a cylindrical lens is commonly placed between the microscope objective and the camera sensor (Cierpka et al. 2010) (see Fig. 1). The induced astigmatism creates two focal planes F_{xz} and F_{yz} which act along the x and y axis of the image, respectively, and generate unique image blurring patterns along the depth axis. For spherical particles, this results in round to star-shaped particle images for particles located at the midpoint between focal planes and vertically or horizontally aligned blurred ellipses for particles located in front or behind the midpoint between focal planes. See Fig. 2 for a depth stack of such particle images or Cierpka et al. (2010) for an illustration of the ray paths in astigmatic arrangements. These differences in defocused images with respect to the optical axis allow for the depth prediction of the particles without changing the camera optics. The resulting optical setup is ideal for tracking tracer particles in small flowing systems. Unlike stereo-methods (Holzner et al. 2015), this technique requires optical access only from one angle, allowing for the use of microscopes as a platform for the investigation of particle movements in small volumes. In such an astigmatic optical system, image acquisition speed is only limited by the frame rate of the camera. However, there are limitations to this approach which stem from current image analysis methods that are focusing on low particle image densities.

Several techniques exist which relate defocused images of spherical tracer particles to their three-dimensional location in astigmatic optical systems. All of these methods consist of two separate steps: First, particle images are detected and image coordinates of the particle image center

are determined. In a second stage, the depth coordinate of particles is inferred from the defocused particle images.

Most methods employ a grayscale threshold after a combination of smoothing and filtering operations for particle image detection. Remaining connected pixels are assumed to belong to the same particle image (Chen et al. 2009; Cierpka et al. 2010; Barnkob et al. 2015). After thresholding, a cascade of cross-correlation steps can be applied to probe the blobs of connected pixels for overlapping individual particle images (Lei et al. 2012). This procedure was developed for particle image detection of round, Gaussian shapes. This occurs in particles which are close to the focal plane in conventional optical arrangements and is therefore not applicable for large depth measurements where the image shape is changing. After detection of particle images, the in-plane locations are determined by fitting Gaussian or generalized Gaussian distributions to the individual particle images (Chen et al. 2009; Cierpka et al. 2010; Barnkob et al. 2015; Lei et al. 2012; Franchini et al. 2019).

Current methods for decoding the depth position from astigmatic defocused images can be grouped in two types. The first type uses the previously determined fitting parameters of the distributions as features to infer the depth coordinate (Chen et al. 2009; Cierpka et al. 2010, 2011; Franchini et al. 2019). The second type derives the depth location by applying cross-correlation between a defocused particle image and the particle images in a lookup set of images with known depth locations (Barnkob et al. 2015). The depth location of the image in the lookup dataset with the maximum cross-correlation value is assigned to the particle of investigation.

Particle localization in images of high particle image density is particularly favorable in the field of microfluidics. The higher the density of localized and tracked particles, the lower the error of spatially interpolated quantities such as the velocity field. Higher particle densities result in larger numbers of overlapping particle images. Yet, both detection and depth prediction stages of current methods are geared towards the detection of non-overlapping particle images.

On the detection stage, a simple filtering and thresholding detection might not detect overlapping particle images if the centers of particle images are proximal and do not form two distinctive intensity peaks or the threshold value is set below the value which allows for detection of individual peaks. Such overlapping instances are simply filtered out in the depth prediction step (Cierpka et al. 2010, 2011; Barnkob et al. 2015).

To allow for the separate detection of two particle images which are overlapping and still form distinctive peaks, the threshold would have to be high enough to result in separate connected pixel regions for each particle. The closer the particle centers are to each other, the higher the intensity threshold which leads to two separate

Fig. 2 Smoothed stack of defocused images of a single spherical tracer particle along the depth axis. The darker the color, the higher the drop in pixel intensity caused by the shadow of the particle falling on the camera sensor under bright-field illumination. Negative values on the depth axis point towards the camera. The unique particle images along the depth axis are a result of the focal planes F_{xz} and F_{yz} , which act on image features oriented along the horizontal (X) and vertical (Y) image axis, respectively

connected pixel regions. However, such high thresholds would result in undetected particle images with intensity maxima below the threshold value. Thus, applying a global threshold for particle image detection results in a trade-off between the detection of low-intensity particle images and the separate detection of particle images with superimposed particle images.

In the case where partially overlapping particle images are detected, image features can be extracted by co-optimization of generalized Gaussian distributions for all overlapping particle images (Franchini et al. 2019). This achieves depth localization of partially overlapping particle images at a high computational cost. The Levenberg–Marquardt method is frequently employed to solve such nonlinear fitting procedures. The computational cost of the method is (Lin et al. 2016):

$$\text{cost} \approx \mathcal{O}(m \times n^2) + \mathcal{O}(m \times n) + \mathcal{O}(n^2) \quad (1)$$

where m is the number of pixels of the image region with the overlapping particle images and n are the number of fitting parameters of a Gaussian or a generalized Gaussian distribution. When extracting image features from multiple particle images simultaneously, the number of fitting parameters is proportional to the number of particle images. Hence, the run time complexity is $\mathcal{O}(N^2)$ and the method becomes less efficient with an increasing number of particle image overlaps.

Thus, current methods are limited in their simultaneous detection of overlapping particle images and low-intensity particle images and their computational cost might increase when processing overlapping particle images. Yet, in fluid flow experiments it is often desirable to conduct dense measurements to fully resolve flow phenomena. Therefore, we focus on three challenges when analyzing astigmatic defocused images. We propose a method which is able to

- detect overlapping particle images even in cases where two particle images share the same center locations
- has a computational cost which is independent of the number of particle image overlaps at inference time
- detects low-intensity particle images close to the human visibility limit

For simultaneous detection of overlapping and low-intensity particle images, we employ an end-to-end trainable neural network. It comprises a convolutional neural network for the encoding of image features and a long short-term memory network for decoding the particle locations (Stewart et al. 2016). The training of the network requires a set of fully annotated images of our application. This means that we need a dataset comprising images where the 3D position of all imaged particles is known.

In the first part of this work, we demonstrate how we created a fully annotated set of images in a cut-and-paste fashion (Dwibedi et al. 2017). The resulting semi-synthetic set of images comprises real particle images while we have complete control over the particle image density and the proximity between particle image center locations. We go on to explaining the architecture of the employed model, our modifications and how the model was trained. The uncertainty of the model is then shown with regard to the model performance in semi-synthetic images as well as real video sequences of moving tracer particles entrained in a flow field in a mini channel.

2 Methods

2.1 Generation of semi-synthetic dataset

2.1.1 Image acquisition and processing

The tracked particles in our experiments are polystyrene microspheres of 2 μm diameter which are dispersed in glycerol and imaged in a rectangular borosilicate channel (CM Scientific) of 3 mm depth, 7 mm width and 30 mm length (see Fig. 1). The glycerol is filtered, the rectangular channel is cleaned, and the refractive indices are matched such that no other objects apart from the microspheres are visible. The experimental scene is imaged with an inverse microscope (Leica DMi8) with 5 $\times$ infinity corrected objective (Leica HC PL FLUOTAR 5) and 2.5 $\times$ camera adapter (Leica 10446175). A cylindrical lens (Thorlabs plano-convex round cylindrical lens, $f = 250$ mm) is placed inside the housing of the camera adapter at a distance of approximately 185 mm to the camera sensor to induce the optical astigmatism into the system. Images were acquired under bright-field illumination and captured with a camera with CMOS sensor (Phantom Miro Lab 340) at a resolution of 1440 $\times$ 2560 pixels with a frame rate of 24 frames per second and an exposure time of 550 μs . With conventional bright-field illumination from the standard LED light source of the transmitted light arm of the microscope (Leica 11525116), the setup produces images with a signal-to-noise ratio of 5.6. The signal is defined as the difference between the minimum pixel value of a particle image and the mean background intensity. The noise is the standard deviation of the background.

Images of the microspheres are visible for up to approximately 200 μm on either side of the midpoint between focal planes before they start fading into the background noise. The resulting dimensions of the imaged volume are 2048, 1152 and 400 μm in length, width and depth, respectively.

As a first step in the creation of a semi-synthetic dataset, we relate a large number of particle images to their three-dimensional location. We annotate images of 380 particles

to account for all variations in particle images. Such variations stem from slight differences in particle shape as well as spherical aberration depending on the particle location in the image plane. It is important to include all variations of particle images into this dataset, as the dataset is later used to train a machine learning model. Missing variations of particle images might lead to poor model performance. To include these variations in particle images, we image all tracer particles along the entire depth of the experimental scene. An image sequence is acquired which spans the depth of the rectangular channel by moving the focal planes at a fixed distance of $3\text{ }\mu\text{m}$ from bottom to top with the focal drive of the microscope. Thus, a focal drive, available for instance for standard inverse microscopes, is necessary for this volume scanning approach. With this procedure, each particle is imaged along the entire depth in which it is visible. This amounts to approximately 120 focal positions at which each particle is imaged. At each location, 4 images are acquired of exactly the same particle distribution. Automatic steering of the focal drive and triggering of the camera was conducted with standard microscope software. The additional images of the same scene are important later on in reducing the noise level of individual particle images. Due to the high viscosity of the glycerol, the particles can be considered stationary during this process.

Possible in-homogeneous background illumination and differences in pixel sensitivity can be homogenized for each pixel on the sensor. Such a pixel-wise correction of the background illumination and the scaling between darkest and brightest possible grayscale value can be achieved by adjusting the flat-field correction (Seibert et al. 1998) to our purpose for bright-field images:

$$I_{\text{corr}} = 1 - \frac{I_{\text{raw}} - I_{\text{dark}}}{s \times (I_{\text{bright}} - I_{\text{dark}})} \quad (2)$$

where I_{corr} is a flat-field corrected image with dark background and bright particle images, I_{raw} is a raw image, I_{dark} is an image of the darkest possible pixel values, I_{bright} is an image of the brightest possible pixel values, and s is a parameter which scales the gain of the image. I_{bright} is an image which is acquired under experimental background light conditions but without any particles in the flow channel. I_{dark} can be acquired by simply lowering the light intensity of the background illumination. This scales between the brightest and darkest possible value that each pixel of the camera sensor can experience.

2.1.2 X, Y, Z annotation

The centers of non-overlapping particle images are estimated by visual inspection in every 10th image of the previously described depth image stack. This form of manual

Fig. 3 Top left: particle image in the center of cropped square region of 70×70 pixel. Top right: signal mask (magenta) and background mask (green). Bottom left: signal mask applied to cropped particle region. Bottom right: background mask applied to cropped particle region

annotation was chosen to ensure that no overlapping particle images were included in the annotated dataset. The estimated center points are refined by fitting a 2D generalized Gaussian distribution to particle images:

$$f(x, y) = \alpha \times \exp \left[\frac{|x_{r, \text{cp}} - x_r|^{\beta_x}}{\rho_x} + \frac{|y_{r, \text{cp}} - y_r|^{\beta_y}}{\rho_y} \right] \quad (3)$$

$$\begin{aligned} x_{r, \text{cp}} &= x_{\text{cp}} \times \cos(\phi) - y_{\text{cp}} \times \sin(\phi) \\ y_{r, \text{cp}} &= x_{\text{cp}} \times \sin(\phi) + y_{\text{cp}} \times \cos(\phi) \\ x_r &= x \times \cos(\phi) - y \times \sin(\phi) \\ y_r &= x \times \sin(\phi) + y \times \cos(\phi) \end{aligned}$$

where x and y are the locations of pixels of the particle images, α is the maximum intensity of the particle image, β is the convexity/concavity, ρ is the shape parameter and ϕ is the rotation of particle image due to lateral aberration and x_{cp} and y_{cp} are the refined center points of the particle images.

All refined center-point annotations of each microsphere are connected to sequences of particle images along the depth axis of the experimental scene as follows. The Hungarian algorithm (Kuhn 1955) is employed to connect the refined center points locations which have been determined for every 10th image. The resulting 3D trajectory of each particle is smoothed along the x and y direction and extrapolated by one additional step of 10 images at the top and bottom. The x and y locations are then determined for all images in between the manually annotated images from the smoothed 3D particle trajectories. The smoothing of trajectories improves the precision of the center points estimation

further and yields high-quality annotations of all non-overlapping particle images in each image, including the manually non-annotated images.

These partially annotated images, however, are not used for training of a model, since the overlapping particle images herein are not annotated. Instead we crop all annotated particle images and use them to synthesize a semi-synthetic dataset of images with known three-dimensional location of all particles. The cropped particle images all have the same square dimensions of 70×70 pixel (see Fig. 3 top left for a cropped particle image). The diameters of particle images change drastically depending on the depth position from several pixel to almost 70 pixel. The cropped particle images are then stacked along their entire depth trajectory. We term the resulting depth stacks of particle images “traces.” A trace therefore describes the particle image over the entire volume where it has been manually detected.

The traces of cropped particle images along the depth coordinate of the experimental scene now contain sub-pixel in-plane (x and y) locations of the particles. The depth location of the individual particle images still has to be determined. First, individual particle images are labeled by their index within the trace from lowest to the highest depth position. Due to uncertainty in the detection of far out-of-focus particle images, the lowest depth position of individual traces might not be exactly the same for all traces. The starting depths of the preliminary depth labels are therefore different and have to be shifted with respect to the other traces to align all traces on a common plane. For this purpose, we compute image features which produce characteristic distributions along the depth axis of the traces. Aligning the depth curves of these image features allows for the quantification of the depth shift necessary to align the traces onto a common plane.

Two image features are employed for providing characteristic curves to determine the depth shift of traces. The local variance and a masked signal. We defined the masked signal as the difference between the mean pixel intensity within a signal mask and a background mask (see Fig. 3). The advantage of computing the masked signal in this manner is that it smooths out the intensity peaks of the two individual focal planes and provides one single peak in the distribution (see Fig. 4).

The local variance is computed over the entire cropped rectangular particle image. The closer a particle is to one of the two focal planes, the weaker the defocus blur along the respective image axis and the larger the difference in pixel intensities. This leads to higher values of the local variance for particles close to focal planes, providing a second distinctive function which can be employed to compute the depth shift.

When normalized between values of 0 and 1, both masked signal and local variance of the particle images create a characteristic function with respect to the depth axis. Particles are considered to be depth-aligned when their distributions of masked signal and local variance traces are aligned. We can therefore align traces onto a common plane by shifting them with respect to each other along the depth axis such that their masked signal and local variance functions match. The depth shift of each trace is determined by linearly interpolating the masked signal and local variance measurements and shifting them with respect to the masked signal and local variance measurements of all other traces to minimize the difference between all measured signals and the difference between all local variances. Adding the depth shift to the indices of all particle images in a trace annotates the depth position of all particle images in the trace. When all traces depth-aligned, we add a constant value to the previously

Fig. 4 Left: Normalized masked signal of depth-aligned traces of particle images. Middle: Normalized local variance of depth-aligned traces of particle images. F_{xz} and F_{yz} are the two focal planes acting

on the X and the Y axis of the image. Right: Comparison of distributions fitted to measured masked signal and local variance

determined shifts of all traces to center the zero z -location (see aligned distributions in Fig. 4). This results in the automatic annotation of the depth location of all traces.

The accuracy of this shift-finding method was determined by creating artificial realizations of the masked signal and local variance curves. Fitted Gaussian distributions as shown in Fig. 4 represent interpolated curves of the image features, and the noise can be sampled from the measured deviations from these interpolations. Our shift-finding method is able to determine randomly generated shifts with a standard deviation of 1.04 of the step size of the depth stack. The step size is $3\text{ }\mu\text{m}$ in our dataset.

Local focus functions have previously been employed to locate the focal planes in astigmatic systems (Cierpka et al. 2011). Variance-based methods for the alignment of objects with a focal plane were found to be the most accurate type of focusing algorithm in microscopy (Sun et al. 2004). We are therefore able to determine the two focal planes in the system from the center points of the two Gaussian distributions which were fitted to the measured local variance distribution (see F_{xz} and F_{yz} in Fig. 4).

The resulting set of traces along with their in-plane and depth locations is cleaned in two ways. First, images with a depth location which is lower than the 1st percentile or above the 99th percentile of all locations are removed. This removes depth locations which are sparsely sampled and assumed to not result in robust model predictions. The resulting depth range in which a human annotator was able to detect particle images spans $408\text{ }\mu\text{m}$. Additionally, the dataset is cleaned from particle images which might have features of other particle images in the background. This is achieved by applying a mask for the background of the

image and then determining the variance of the background pixels. Particle images with a background variance above the 95th percentile are removed from the dataset. The result of this annotation workflow is a collection of cropped square images with one three-dimensional-annotated particle images at the center of each image. Each cropped square image was taken four times initially, which will enable us in the next step to fine-tune the noise levels in overlapping sections of the cropped images.

2.1.3 Stitching of overlapping particle images onto synthetic background

To train an algorithm in the detection of overlapping particle images, we require annotated images which contain such overlaps. For this purpose, we generate a semi-synthetic set of training images, containing the annotated cropped images from the previous step, carefully overlapping them and placing them in front of a synthetic background.

First, a set of random locations is generated for placement of a first batch of cropped particle images in the training image (see Fig. 5). This first batch of image locations is sampled at random and large enough distances from each other such that they are not overlapping each other. In a second step, a defined fraction of the first batch of locations are chosen to overlap with at least one more cropped particle image. We chose the fraction of overlapping to be 50% of the number of images of the first batch of cropped particle images. Yet, the actual number of superimposed particle images might be smaller, since the particle images do not fill out the entire cropped image region. However, this sampling scheme ensures that a defined fraction of cropped particle

Fig. 5 Generation of synthetic training images: (1) A first batch of cropped particle images is placed in front of synthetic background noise. The color of the diamond markers indicates their location on the optical axis. (2) A defined fraction of the first-batch cropped particle images are overlapped with a second batch of particle images at a randomly sampled offset. (3) The noise is adjusted in overlap-

ping cropped image areas. (4) Random 2D Gaussian fields are added to simulate background noise from far out-of-focus particle images. (5) Final semi-synthetic image. (6) Close-up of real image. Please note that the intensity colorbar below image 2 is valid for all images except for image 4 which has its own intensity colorbar

images are overlapping and non-overlapping to provide sufficient data for the model to learn the shape of individual particle images as well as to distinguish between overlapping particle images. The employed fraction of cropped particle images with overlap was found to result in an optimal compromise between these two tasks (see Sect. 3.1). It might be possible to improve the performance of one of the tasks on the cost of lower performance on the other task by changing the fraction of cropped particle image overlaps in the training image. However, this was not investigated further in this study.

The locations of the second batch of overlapping cropped particle images are chosen relative to the locations of the cropped particle images of the first batch. Distances in the x and y directions between the center locations of the first and second batch particle images are drawn from a combination of two distributions. A normal distribution with a mean of zero and a variance of the maximum length of particle images accounts for neighboring particle images with a high probability of superimposed particle images. A uniform distribution is added with a total probability of 0.4 to account for neighboring particle images which are close, but do not necessarily overlap. The distance of the second-batch images is limited to half the linear size of the cropped particle images. This process of selecting particle image locations ensures a large degree of overlap, allows for non-overlapping close neighbors and produces a varying particle image density.

The previously determined center point locations are used to place cropped particle images in front of an empty background. For each particle image location, random particle images are sampled from the training set of location-annotated cropped particle images. The mean background value of each particle image is adjusted such that all particle images have the same mean value of the background. In overlapping image areas, grayscale values of two cropped particle images are not simply added up. This would lead to an increase in noise levels as the noise of the individual images would be added. To achieve the same noise level as the non-overlapping cropped particle images, overlapping parts of cropped particle images are first smoothed. This smoothing is achieved by averaging a set number cropped images of the same particle. The number of cropped images which is necessary to achieve the correct noise level in overlapping areas is determined numerically. In our case, 2.3 cropped images of each particle are averaged before the resulting smoothed images are added to produce the image for the overlap area. This results in the same background variance in overlapping areas as compared to non-overlapping non-smoothed images.

The particle images in the training images are based on real images, whereas the image background is filled entirely with synthetic data. The background can be generated by

sampling from a mean background signal and adding a sample of random noise from the camera sensor as well as patches of overlapping, far out-of-focus defocused particle images. Distributions of both mean background signal as well as random sensor noise are measured in real images. By sampling from these distributions, we reproduce the same image conditions as the original experimental images. Defocused patterns in the image background are based on realizations of directional 2D random Gaussian fields. First, we sample random parameters of the random fields and use the same parameters to generate directional Gaussian fields in both the horizontal and vertical direction of the image. We threshold these random fields by a random value to account for image areas which are free of background signals. In a next step, the resulting values of the background pattern are scaled at a random value between zero and the signal of the most out of focus particles in our dataset. Any particles further away from the midpoint between focal planes, which we want to simulate, would thus produce a weaker signal. The synthetic camera noise and the mean background signal are added to the image areas which do not contain any particle images. The defocused background pattern is added to the entire image, to account for the overlap of particle images with undetectable out of focus signals. The resulting semi-synthetic images have in many cases more background patterns than the real experimental images. This is to allow for the training of a model which is robust to low-frequency defocused background patterns.

2.2 Implementation of model for dense particle image detection

For the detection of overlapping particle images, we require an algorithm which is able to distinguish objects of the same class from each other, even if they share the same center point. Most modern object detection algorithms such as SSD (Liu et al. 2016), YOLO (Redmon et al. 2016) and Mask R-CNN (He et al. 2017) output multiple candidate object detections for the same object. Therefore, they rely on non-maximum suppression as a postprocessing step. When candidate object detections of the same object class are overlapping by more than a given threshold, non-maximum suppression only keeps the detection with the highest confidence score and removes all other detections. This renders models which rely on non-maximum suppression unsuitable for our goal of detecting particle images with the same center point location.

Instead, we employed an end-to-end trainable architecture for our purpose (Stewart et al. 2016). This approach uses a convolutional neural network (CNN) to extract image features within regions of the image. The features of each image region are fed separately into a long short-term memory network (LSTM) which translates image features into object

detections. This end-to-end trainable model explicitly learns to produce one detection for each object and thus allows for the detection of arbitrary degrees of object overlap.

Tensorbox is an open-source implementation of the original algorithm. We adapted this model to predict out-of-plane locations of particles in addition to their in-plane locations (see Fig. 6). The model Inception V1 (Szegedy et al. 2015) is employed as the CNN (red bounding box in Fig. 6). In this CNN architecture, the input of the model first passes through several convolutional and pooling layers (yellow and purple boxes in Fig. 6) before inception modules (orange boxes in Fig. 6) are employed instead of the conventional convolutional layers. Inception modules consist of several convolutional and pooling operations which are all applied in parallel to the same input of this layer and are then concatenated together as output. 1×1 convolutional operations in this module reduce the computational budget and allow for a wide and deep neural network architecture. This allows the architecture to learn a wide range of complex image features.

Due to the convolutional and pooling operations, the output of the CNN is three-dimensional with a lower resolution in the image dimensions. The extracted image features are stored in feature vectors along the additional third dimension. The area in the input image that a feature vector is connected to is called the receptive field. In our case, the receptive field of the last inception module of the CNN has a size of 789×789 pixel. Therefore, each feature vector “sees” a large proportion of the input image. This information from a large receptive field is employed to account for large scale defocused background patterns from undetectable particles. However, the region size in which each feature vector is employed to detect particles is only 32×32 pixel.

These feature vectors are fed into a LSTM (Hochreiter and Schmidhuber 1997) with four recurrent units and 800 memory states. The number of memory states determines the number of neurons involved in detecting particles and locating them in 3D based on information from the feature vectors. The number of recurrent units determines the number of outputs of an LSTM and therefore equates to the maximum number of particle image detections in each image region. Each recurrent unit computes horizontal, vertical and depth locations as well as a confidence score c . Linear interpolation of shallower feature vectors at the previously determined particle image locations yields additional information about precise particle locations. These shallow features are derived from the output of the first Inception module of the model and produce an offset vector which updates the particle locations and confidence scores. The feature interpolation step is the only part of the model where the computational cost is proportional to the number of particle images. Yet, it is independent of the number particle image overlaps. The average inference time for an image with 200 particle images was 1.02 s.

The CNN is initialized with weights, which were trained on the Imagenet dataset. The entire network is trained with 10,000 semi-synthetic images, each containing in average 200 particle images from a set of approximately 45,600 individual particle images. Adam (Kingma and Ba 2014) was employed as optimizer and a cyclic learning rate schedule followed by a decaying learning rate. During the initial cyclic period, the learning rate was set to alternate sinusoidal with a frequency of 20,000 parameter updates followed by an exponential decay of the learning rate from parameter update 30,000 to 100,000. We found that this alternation in

Fig. 6 Architecture of modified Tensorbox model: a CNN (red bounding box) acts as feature encoder for image regions. Feature vectors of image regions are fed individually into a LSTM (green bounding box) which creates a sequence of center point locations. These

locations are updated with fine-grained information from the interpolation of early feature vectors at the previously determined in-plane locations of the particle images. The original model was only modified by adding the depth location to the model predictions

the learning rate shortened training time and lowered the final error.

3 Results and discussion

3.1 Results on semi-synthetic dataset

The trained model is first evaluated on a test set of 1000 semi-synthetic images to test the performance of the algorithm for the task of inferring the depth position from defocused particle images. The model predictions indicate that the model architecture is capable of simultaneously decoding both in-plane and depth positions from defocused

particle images. Figure 7 shows the relationship between depth prediction and the true depth values. There is no visible bias in the depth inference along the depth axis, nor is there a visible deterioration in accuracy with distance from the midpoint between focal planes. The algorithm is capable of inferring particle depth positions up until the visibility limit and covers the same depth range as the human annotator of 408 μm . The model has a slightly higher variance around the regions of the two focal planes.

At the visibility limit of the defocused particle images at a distance of approximately 200 μm from the midpoint between focal planes, the depth location error again increases. When the detected particle images are filtered by their confidence score, it can be seen that low-confidence scores are clearly related to the particle images at the fringe of their visibility. This is where their signal is getting close to the signal of the artificial defocused background patterns. The defocused particle images are starting to merge into the image background which might be the cause of the lower confidence scores.

The overall rate of undetected particle images (false-negative detections) was found to be 2.6%, and the rate of false-positive detections was found to be 2.7%. Undetected particle images are clearly related to the presence of particle image overlaps (see Fig. 8). The closer a neighboring particle image is to the center location of a particle image, the higher the rate of undetected particle images. The proximity of false-positive detections to real particle images reveals that most false-positive detections correspond to particle images which are interpreted as two particle images. The equal performance with regard to false-positive and false-negative detections indicates that the initially chosen fraction of 50% cropped particle image overlap resulted in a

Fig. 7 Estimated depth position versus true depth position of all detected particle images in the validation set of the semi-synthetic dataset. F_{xz} and F_{yz} are the two focal planes acting on the X and the Y axis of the image

Fig. 8 Top row: effect of relative location of closest neighbor on detection rate in synthetic images. Rate of undetected particle images (column 3) is equal to the share of the local distribution of undetected particle images (column 2) with respect to the true distribution of the relative location of closest neighbors (column 1). Relative positions

of false-positive detections (column 4) are mainly within a 4 pixel radius around a real particle image. Errors were evaluated at a resolution of 1 pixel. In white: dimensions of the cropped particle images. Bottom row: cross sections of the above plots along $X = 0$ and $Y = 0$. 1 μm corresponds to 1.25 pixel

well-balanced performance of both the detection of single particle images and the distinction of overlapping particle images.

The objective might change for experiments with significantly larger or smaller particle image densities which would result in different numbers of overlaps. For high particle image densities, it could be more important to distinguish between overlapping particle images or for smaller particle image densities it could be more important to reduce false-positive predictions. It might be possible to address such a change in the objective by changing the ratio of cropped particle image overlaps in the training set. This might shift the attention of the algorithm during model training towards the one or the other task. However, such optimization of the proposed technique to different experimental conditions such as particle image density still requires further investigation.

In our semi-synthetic dataset, we are able to compute the true distribution of the relative locations of the closest neighboring particle image. The share of undetected particle images to this distribution provides us with the rate of undetected particle images as a function of the relative location of the closest neighbor. Again, the maximum of this rate is located at the points where particle images share the same center point locations. Here, the rate of undetected particle images is 20.7%. At a center-point distance of 10 pixels and 15 pixels, this value drops to 7.8 and 1.3%. False-positive detections are mainly located within a radius of 4 pixels of a real particle image center.

We further analyzed the impact of neighboring particle image proximity on the performance of the algorithm to infer in-plane and depth locations. For this purpose, we filtered all detected particle images with respect to their distance to their nearest neighboring particle image. Figure 9 shows the location error in the image plane and along

the depth axis as a function of the location of the nearest neighbor. The average in-plane location errors of particle images without overlap are 0.65 and 0.57 μm in the x and y directions, respectively. Location errors start increasing sharply if neighboring particle images are within a distance of less than approximately 10 pixels. When particle image centers share the same in-plane location, x and y location errors almost triple. This effect is even more pronounced for the depth location error. The average depth location error is 6.34 μm for particle images without overlap. When the center points of overlapping particle images are the same, the depth location error increases by a factor of 6. Thus, particle image overlap has a stronger relative impact on the performance of determining the depth position than the in-plane locations. It appears that particle image features which relate to the in-plane locations are in comparison easier for the algorithm to decode during particle image overlap as compared to particle image features which relate to the depth position. Both x and y location errors are slightly higher for neighboring particle image locations along the x and y image axis, respectively. This could be explained with scenes where superimposed particle images are on the same side of the midpoint between focal planes and their elliptical shapes line up on the x or y axis. In this arrangement, their signals are more convoluted and the solution for separating the two signals might become less unique. The total location error is entirely dominated by the location error of the depth error. The sum of the x and y location errors only contributes to approximately 10% of the total location error. It is in accordance with previous works that the depth error is an order of magnitude higher than the in-plane errors (Barnkob et al. 2015; Franchini et al. 2019).

Fig. 9 Top row: location errors of model predictions in images of validation set of semi-synthetic dataset with respect to the relative location of the closest neighboring particle image center. Errors

were evaluated at a resolution of 1 pixel. In white: dimensions of the cropped particle images. Bottom row: cross sections of the above plots along $X = 0$ and $Y = 0$. 1 μm corresponds to 1.25 pixel

3.2 Results on real images

To test the resemblance of the semi-synthetic dataset with real images, we analyzed the performance of the trained model on videos of real images with moving spherical microspheres. We acquired an additional dataset based on the previously described experimental setup with the same fluid, particles and particle density. Image sequences of the particles are captured at 21 different depth locations $150\text{ }\mu\text{m}$ apart from each other throughout the entire depth of the channel. At each depth location, a sequence of 452 images was acquired at a frame rate of 24 fps. Polystyrene tracer particles were added to the glycerol at a concentration of $5 \times 10\text{ w/v}$. Approximately 200 particle images are detectable with human vision on the 1440×2560 pixel camera images. The resulting particle image density of 5.4×10^{-5} particles per pixel is the same as in the semi-synthetic training images. The particle image diameters in the x and y directions vary between approximately 7 and almost 70 pixel depending on the depth position of the particle.

In this experiment, we applied a constant flow rate of 2.3 ml/h to the liquid which is carrying the tracer particles. The liquid is flowing in the rectangular channel approximately parallel to the midpoint between focal planes. The constant flow rate ensures that the particles are moving in a laminar flow on straight trajectories. Therefore, the three-dimensional displacements of particles are constant over time in this flow field.

These constant displacements are an ideal basis for evaluating the variance of the predictions of our trained model. We employ the model on all acquired image sequences and link the detected 3D locations of the particles with the Hungarian algorithm. This algorithm only links particle detections between consecutive frames. More advanced tracking algorithms take information about the flow field into consideration and might perform better in more complex flows (Pereira et al. 2006). All resulting trajectories below a length

of 100 frames were discarded to ensure that trajectory-wise measured error quantities are well-estimated. 38.3% of all recovered trajectories contained a particle image overlap at some point in time. Due to the varying particle image shapes, we defined a particle image overlap as a neighboring particle image center being closer than 35 pixel along the x or y axis to another particle image center. Thus, particle images might not always be superimposed in the fraction of trajectories which were identified as containing a particle overlap at some point in time.

Some trajectories were a result of false-positive predictions. In such cases, one particle was consistently detected twice and can be filtered out. To test whether two close trajectories are the result of different physical particles, the average relative displacements between the two trajectories are compared. Two distinctive particles travel at different velocities if they are located at different depths. Double predictions of the same particle can only move at the same velocity since they are mostly located within a 5 pixel radius (see Fig. 8). Thus, false-positive trajectories can be filtered out by setting a minimum value for the average relative displacement between a particle and its closest neighbor. The threshold is only applied to trajectories with neighbors which are closer than a certain limit. This process efficiently filters out cases where a particle is consistently detected twice. This filtering approach removes particles with accompanied false-positive detections which both have the same trajectories along the x - y plane. Therefore, we expect this filtering procedure to be robust in more complex flows too, where trajectories along the x - y plane differ even stronger than in this study.

Individual displacements along the X axis were derived from all trajectories and visualized (see Fig. 10, left). The displacements in the direction of the X axis do not vary significantly along the Y axis. This is due to the fact that only the centermost $1152\text{ }\mu\text{m}$ of a 7-mm-wide channel has been investigated. The effects of the walls along the Y axis on

Fig. 10 Left: Z - Y profile of measured individual displacements along X axis of the flow channel. The measurement volume covers the central $1152\text{ }\mu\text{m}$ of the 7-mm-wide channel. Right: depth profile of measured individual displacements along the X axis compared to analytical solutions of the flow field. Analytical solutions were computed for

the center and side of the measurement volume (i.e., at $Y = 0\text{ }\mu\text{m}$ and $Y = 576\text{ }\mu\text{m}$). Note that the variance of displacements increases at the top and bottom of individual measurement slices, visualized in more detail in Fig. 11. $1\text{ }\mu\text{m}$ corresponds to 1.25 pixel, yielding a maximum velocity of the analytical solution of $49.7\text{ }\mu\text{m s}^{-1}$

the flow field are therefore not prominent in the measured volume. The periodic change in displacement variance along the depth axis (Fig. 10, right) shows that within each individual measurement slice, the error of the measured displacement changes significantly along the depth axis.

To estimate displacement errors, we first approximated the constant displacements of particles by calculating the median displacements along trajectories. The residuals between those averages and the individual measured displacements allow us to study the variance of the model predictions from real images without the need for a fully annotated dataset. We did not observe any significant bias in the location errors of our model with respect to the semi-synthetic dataset (see for instance Fig. 7 for depth error). Thus, we assume that the variance of the model predictions from the real images is a good approximation for the total location error (Cierpka et al. 2011).

Figure 11 shows how location errors of each trajectory relate to the average depth prediction of the particle. The in-plane location errors in the x and y direction show a very similar behavior. Close to the midpoint between focal planes, their error reaches a minimum of approximately $0.3 \mu\text{m}$, and towards the upper and lower limit of particle detectability, the error increases up to approximately ten times higher values. It can be seen that the x error increases more strongly towards small depth values and that the y error increases more strongly towards large depth values. This can be explained with the shape of the particle images at those depth positions. For small depth values, the image of a microsphere forms an ellipse, with its long axis aligned along the x axis (see Fig. 2). For such a particle image, the image gradient is larger along the y axis than the x axis. It is therefore easier to locate the center location of the particle image along the y axis than the x axis for low depth values. The opposite is true for large depth values, where the long

axis of the particle image is aligned with the y axis of the image and the y location error is larger.

Previously reported in-plane location errors from the fitting of 2D generalized Gaussian distributions to particle images are 0.21 pixels for images with the same signal-to-noise ratios (Franchini et al. 2019). This refers to non-overlapping and “slightly” overlapping particle images. The error corresponds to $0.17 \mu\text{m}$ by taking the magnification of our apparatus into account. Our minimum in-plane errors of $0.3 \mu\text{m}$ are therefore 76% larger as compared to the current method of locating tracer particles in the image plane. The sum of the differences between in-plane location errors of the two methods contributes to approximately 2% of the average total location error. Depending on the application, this additional error might not be significant for a user.

The depth location error in our study shows a less clear trend along depth axis. It reaches minimum values of approximately $6 \mu\text{m}$ where the masked signal reaches its maximum value and there are several regions with local peaks in depth error. Focal plane F_{yz} correlates with such a local error peak. Focal plane F_{xz} corresponds to higher error values; however, it does not fall directly on local error peak. This is consistent with the performance of the algorithm of Cierpka et al. (2011), which showed a local error peak close to one focal plane and another local error peak in the vicinity of the second focal plane. Maximum displacement error values of approximately $15 \mu\text{m}$ are reached at the detectability limit. The relative increase in error with distance to the depth center of the measurement volume is therefore not as strong as in the case of the in-plane location errors. Yet, the number of outliers which experience larger errors is significantly larger for the depth location error.

Previous studies reported similar ratios of in-plane to depth location errors with ratios ranging between 1:10 and 1:30 (Cierpka et al. 2010, 2011; Barnkob et al. 2015;

Fig. 11 Displacement errors in X direction versus depth location. X and Y are the in-plane coordinates, and Z depicts the depth coordinate. Each point represents the standard deviation of the residuals between individual displacements and the median displacement of one trajectory in the respective directions. The true standard deviation of displacement errors is better estimated with longer trajectories. To

visualize this higher significance of longer traces, the sizes of measurement points reflect the trajectory length as shown in the legend of the left plot. F_{xz} and F_{yz} are the two focal planes acting on the X and the Y axis of the image and $S_{\max}$ is the location of the maximum masked signal. $1 \mu\text{m}$ corresponds to 1.25 pixel

Franchini et al. 2019). Location errors strongly depend on the signal-to-noise ratio of the experimental apparatus (Cierpka et al. 2011). A study with images of the same signal-to-noise ratio reported a depth location error of $7.1\text{ }\mu\text{m}$ with regard to their calibration dataset of non-overlapping particle images (Franchini et al. 2019). Their calibration dataset serves the same purpose as our semi-synthetic dataset: to train a model to predict particle locations. In our study, we achieved a depth location error of $6.3\text{ }\mu\text{m}$ in our synthetic dataset. The average depth error of non-overlapping trajectories in our real images is $9.2\text{ }\mu\text{m}$ over a 67% larger depth range than (Franchini et al. 2019). The difference in the errors between the semi-synthetic dataset and the real dataset of our study might be explained with a slight difference in the distributions of the data in the two dataset. The semi-synthetic dataset might include image features which are not present in the real data and vice versa. The same might be true for the stated depth location error with respect to the calibration dataset in Franchini et al. (2019). We therefore consider our measured depth location error to be of a comparable level of accuracy.

The depth range in which the trained algorithm is able to recover particle trajectories spans $373\text{ }\mu\text{m}$ (see Fig. 11). This covers 91.4% of the depth range in which a human annotator was able to detect particle images in real experimental images. This translates into an increase in the depth range along which particles can be detected of 67 % as compared to detection by simple smoothing and thresholding (Franchini et al. 2019). This indicates that the semi-synthetic dataset was successful in representing the differences between defocused background patterns and low-intensity particle images. The performance of the employed neural network is therefore close to human performance in detecting low-signal particle images.

Location errors as a function of the closest neighboring particle image show a similar behavior as in the synthetic dataset (see Figs. 9, 12); however, there is clearly more noise in the error values of the experimental dataset. This is most

likely due to a smaller number of sample points. For all location errors, there is a clear trend towards higher errors for distances to the nearest neighbor of less than approximately 10 pixel. The error values at larger distances to the particle image center point are comparable to the values in the synthetic dataset. In this region, average in-plane location errors range between 0.5 and $1.0\text{ }\mu\text{m}$, whereas z -location errors range between 6 and $11\text{ }\mu\text{m}$. The total location error approximately doubles to a maximum total location of $14.7\text{ }\mu\text{m}$ when two neighboring particle images share the same center point. For neighbor locations which share the same particle image center, all location errors are below the errors of the synthetic dataset. This shows that the model is capable of detecting overlapping particle images. Yet, more challenging particle image overlaps which caused higher errors in the synthetic dataset are possibly not detected in the real images.

4 Conclusion

In this work, we showed how a set of semi-synthetic images can be employed to train a neural network in the task of detecting overlapping defocused images of tracer particles. The semi-synthetic images, which we generated for model training, resemble experimentally acquired images. The model is able to detect 97.4% of particle images in the semi-synthetic images, despite 50% of particle images containing a particle image overlap with another particle image. Due to variable particle image shapes, particle image overlaps were defined as two particle center points being located in a square image region of 35×35 pixel. Particle images might not always be overlapping for this definition of a particle overlap. However, when two defocused particle images share the exact same center point location, 79.3% of particle images are still detected. The share of false-positive detections in this set of images is 2.7%.

The trained model generalizes well to experimental images. It increases the depth range at which particles are

Fig. 12 Location error defined as measured residuals between median trajectory displacements and actual displacements as a function of the relative in-plane location of the center of the closest particle image. X

and Y are the in-plane coordinates and Z depicts the depth coordinate. In white: dimensions of the square particle images. $1\text{ }\mu\text{m}$ corresponds to 1.25 pixel

detectable by 67% over a conventional method of thresholding grayscale values by Franchini et al. (2019). The resulting depth range of 373 μm is 91.4% of the depth range in which a human annotator was able to detect particle images. Particle image overlap impacts the location errors of detected particles. Overlapping particle images with a common center point cause the total location error to increase by approximately a factor of two as compared to non-overlapping particle images.

We hope that this work opens up new applications of defocused particle image localization where large depth ranges or higher particle image densities are required. Our application of this method will be the study of 3D flow in porous media with particle tracking. Yet, others might find it useful to observe the movement of microorganisms or to determine velocity fields for the investigation of heat and mass transfer.

Acknowledgements We gratefully acknowledge funding from the Qatar Carbonates and Carbon Storage Research Centre (QCCSRC), provided jointly by Qatar Petroleum, Shell, and the Qatar Science and Technology Park.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barnkob R, Kähler CJ, Rossi M (2015) General defocusing particle tracking. *Lab Chip* 15(17):3556–3560
- Chen S, Angarita-Jaimes N, Angarita-Jaimes D, Pelc B, Greenaway A, Towers C, Lin D, Towers D (2009) Wavefront sensing for three-component three-dimensional flow velocimetry in microfluidics. *Exp Fluids* 47(4–5):849
- Cierpka C, Segura R, Hain R, Kähler CJ (2010) A simple single camera 3C3D velocity measurement technique without errors due to depth of correlation and spatial averaging for microfluidics. *Meas Sci Technol* 21(4):045401
- Cierpka C, Rossi M, Segura R, Kähler C (2011) On the calibration of astigmatism particle tracking velocimetry for microflows. *Meas Sci Technol* 22(1):015401
- Dwibedi D, Misra I, Hebert M (2017) Cut, paste and learn: surprisingly easy synthesis for instance detection. In: *Proceedings of the IEEE international conference on computer vision*, pp 1301–1310
- Franchini S, Charogiannis A, Markides CN, Blunt MJ, Krevor S (2019) Calibration of astigmatic particle tracking velocimetry based on generalized Gaussian feature extraction. *Adv Water Resour* 124:1–8
- Guo Q, Alexander E, Zickler T (2017) Focal track: depth and accommodation with oscillating lens deformation. In: *Proceedings of the IEEE international conference on computer vision*, pp 966–974
- He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, pp 2961–2969
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Holzner M, Morales VL, Willmann M, Dentz M (2015) Intermittent Lagrangian velocities and accelerations in three-dimensional porous medium flow. *Phys Rev E* 92(1):013015
- Kingma DP, Ba J (2014) Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*
- Kuhn HW (1955) The Hungarian method for the assignment problem. *Naval Res Logist Q* 2(1–2):83–97
- Lei YC, Tien WH, Duncan J, Paul M, Ponchaut N, Mouton C, Dabiri D, Rösgen T, Hove J (2012) A vision-based hybrid particle tracking velocimetry (PTV) technique using a modified cascade correlation peak-finding method. *Exp Fluids* 53(5):1251–1268
- Lin Y, O'Malley D, Vesselinov VV (2016) A computationally efficient parallel Levenberg–Marquardt algorithm for highly parameterized inverse model analyses. *Water Resour Res* 52(9):6948–6977
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC (2016) SSD: single shot multibox detector. In: *European conference on computer vision*. Springer, pp 21–37
- Pereira F, Stürer H, Graff EC, Gharib M (2006) Two-frame 3D particle tracking. *Meas Sci Technol* 17(7):1680
- Raffel M, Willert CE, Scarano F, Kähler CJ, Wereley ST, Kompenhans J (2018) *Particle image velocimetry: a practical guide*. Springer, Berlin
- Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 779–788
- Seibert JA, Boone JM, Lindfors KK (1998) Flat-field correction technique for digital detectors. In: *Medical imaging 1998: physics of medical imaging*. International Society for Optics and Photonics, vol 3336, pp 348–355
- Stewart R, Andriluka M, Ng AY (2016) End-to-end people detection in crowded scenes. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 2325–2333
- Subbarao M, Surya G (1994) Depth from defocus: a spatial domain approach. *Int J Comput Vis* 13(3):271–294
- Sun Y, Duthaler S, Nelson BJ (2004) Autofocusing in computer microscopy: selecting the optimal focus algorithm. *Microsc Res Tech* 65(3):139–149
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1–9
- Watanabe M, Nayar SK (1998) Rational filters for passive depth from defocus. *Int J Comput Vis* 27(3):203–225
- Wu M, Roberts JW, Buckley M (2005) Three-dimensional fluorescent particle tracking at micron-scale using a single camera. *Exp Fluids* 38(4):461–465

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.