Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Defocus particle tracking: A comparison of methods based on model functions, cross-correlation, and neural networks

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

This version is available at: https://hdl.handle.net/11585/918481 since: 2024-06-17

Published:

(Article begins on next page)

Defocus particle tracking: A comparison of methods based on model functions, cross-correlation, and neural networks

Rune Barnkob[1],[*] Christian Cierpka[2],[†] Minqian Chen[3], Sebastian Sachs[2], Patrick Mäder[3], and Massimiliano Rossi[4][‡]

[1]*Heinz-Nixdorf-Chair of Biomedical Electronics, Department of Electrical and Computer Engineering,*
Technical University of Munich, TranslaTUM, 81675 Munich, Germany
[2]*Institute of Thermodynamics and Fluid Mechanics, Department of Mechanical Engineering,*
Technische Universität Ilmenau, P.O. Box 100565, D-98684 Ilmenau, Germany
[3]*Group for Software Engineering for Safety-Critical Systems,*
Technische Universität Ilmenau, P.O. Box 100565, D-98684 Ilmenau, Germany
[4]*Department of Physics, Technical University of Denmark,*
DTU Physics Building 309, DK-2800 Kongens Lyngby, Denmark

Defocus particle tracking (DPT) has gained increasing importance for its use to determine particle trajectories in all three dimensions with a single-camera system, as typical for a standard microscope, the workhorse of today's ongoing biomedical revolution. DPT methods derive the depth coordinates of particle images from the different defocusing patterns that they show when observed in a volume much larger than the respective depth of field, therefore it has become common for state-of-the-art methods to apply image recognition techniques. Two of the most commonly and widely used DPT approaches is the application of (astigmatism) particle image model functions (MF) and normalized cross-correlations between measured particle images and reference templates (CC). Though still young in the field, the use of neural networks (NN) is expected to play a significant role in future and more complex defocus tracking applications. To assess the different strengths of such MF, CC, and NN defocus tracking approaches, we present in this work a general and objective assessment of their performances when applied to synthetic and experimental images of different degrees of astigmatism, noise levels, and particle image overlapping. We show that MF-based approaches work very well in low-concentration case, while CC-based approaches are more robust and provide better performance in cases of larger particle concentration and thus stronger particle image overlap. The tested NN-based methods generally showed the lowest performance, however, in comparison to the MF- and CC-based methods, they are yet in an early stage and have still great potential to develop within the field of defocus particle tracking.

I. INTRODUCTION

Several methods have been proposed to enable the fully three-dimensional measurement of the velocity field in a volume. Especially in the case of investigating microfluidic flows, usually only one direction of observation is available, limiting the range of applicable techniques [8]. On the other hand, flows in microfluidics are often inherently three-dimensional or show strong gradients in all directions of the velocity field. One of the first applications using defocused particle images to determine out-of-plane velocities was shown in 1994 by Stolz and Köhler [41]. Since then, several different defocusing-based methods have widely been used, especially in microfluidic applications [4–7, 18, 25–27, 30, 35, 39, 43, 44]. Each of the methods performed well for a specific experiment but a general assessment on the applicability for different situations and a thorough and comprehensive discussion on errors is missing so far.

With one of its first introductions by Kao and Verkman in 1994 [17], one widely-used method applies cylindrical lenses to introduce astigmatic aberrations to break

the axis symmetry of the optical system and to provide elliptically-shaped particle images. By the evaluation of the length and width of the elliptical particle images, a particle's depth position can be determined [11]. However, the algorithms developed for the classical implementation of astigmatism particle tracking velocimetry (APTV) rely on the model assumption of Gaussian-distributed light intensity of the particle images [9]. Although the method showed better performance in comparison to typically-used stereoscopic micro Particle Image Velocimetry (μPIV) [10], there was much room for further improvements and optimization [36]. Especially, since optical systems are quite complex and additional image aberrations often occur, the elliptical model often does not hold and experimentally-acquired images are needed for calibration [9].

Derived from the need of calibration images and to account for differently-distributed intensities, a direct comparison between measured and calibration particles images was proposed, called the General Defocusing Particle Tracking (GDPT) [1]. In its simplest form GDPT, uses templates of actual calibration particle images to determine a particle's depth position through matching of a template particle image and an actual measured particle image using cross correlation. The advantage is that the method can be applied even using minor quality optical systems without knowledge of the transfer function. Additionally, the use of calibration images enables the pre-

measurement assessment of expected uncertainties which can be a critical factor in the adjustment and optimization of a given defocus tracking setup.

Different image comparison approaches can be used and, only recently, artificial neural networks were applied for image recognition and classification in defocus particle tracking problems. These methods have shown to be useful for the position recognition of particle images [12, 21].

The three methods all rely or benefit strongly from the use of calibrations images, each with their advantages in terms of ease of applicability, coordinate uncertainties, detection rates, and required amount of calibration/training data. The aim of the paper at hand is therefore to set up and perform tests of the three different approaches to determine advantages and limitations of each of them in order to guide a potential user to select the best suitable approach for a given application. For this reason and similar to the assessment scheme already used for GDPT [3], we present and apply the three methods to synthetic test images taking into account different degrees of astigmatic aberrations, noise levels, and particle image concentrations [33]. For reference, the methods are further applied to experimental test images. All datasets presented will be freely available to the research community through the community platform https://defocustracking.com/.

II. METHODS

Three different approaches to determine the depth position by certain features of the particle images are compared in this study. The first method relies on some knowledge of the imaging and makes assumptions and simplifications of the particle images, i.e. the spatial intensity distribution on the sensor. The original APTV approach makes use of a model for the width and height of the particles images and uses curve fitting for calibration as will be described in Section II A.

If such a model is unknown or too complex, particle images from known depth position can be used as templates for a cross-correlations with particles images from particles at unknown depth positions. The depth position can then be determined by the template with the highest correlation coefficient as outlined in section Section II B.

A third method is the use of image recognition by modern machine learning neural networks. In principle every image aberration can be regarded as a feature to determine the depth position with such a network if trained by enough particle images where the depth positions are known. The details of such models are described in section Section II C.

All methods have in general that calibration images with known z-positions have to be available to built a calibration function or to train the neural network in order to relate certain image features with the depth position.

These calibration images can be obtained with a synthetic image generator which is especially useful to analyze the performance of the different methods and will be described in Section III A. However, in a real experiment particles, typically dried on a microscopic slide or settled at the bottom of a microchannel are imaged for different known z-positions spanning the desired depth range for the measurements as described in Section IV A.

A. Particle image model functions (MF)

Considering geometrical optics, a model for the imaging of small particles was introduced by Olsen and Adrian [29]. Under the assumption that the intensity distribution follows a Gaussian distribution, the size of a particle image can be calculated, taking into account the contributions due to the particle diameter d_p itself (geometric part), diffraction, and defocusing. The particle image diameter on the camera sensor a_I can be modelled as a function of the distance of the particle to the focal plane δz. Assuming the working distance of the lens significantly larger than δz we have [23, 28]

$$
a_I^2\left(\delta z\right) = (Md_\mathrm{p})^2 + 1.49(M\lambda)^2 \left(\frac{n_0^2}{\mathrm{NA}^2} - 1\right) \dots
$$
$$
\dots + (2M\delta z)^2 \left(\frac{n_0^2}{\mathrm{NA}^2} - 1\right)^{-1}, \quad (1)
$$

with λ being the wavelength of the emitted light, n_0 the refractive index of the immersion medium of the lens, and M and NA the magnification and numerical aperture of the lens, respectively. By measuring the particle image diameter and solving Eq. (1) for δz, it is possible to derive the depth position of the particle with respect to the focal plane, although without knowing its sign. APTV methods solve this problem by introducing a cylindrical lens in the optical path. When properly aligned, it produced elliptical particle images and by measuring the particle image diameters in the horizontal and vertical direction (a_x and a_y) it is possible to derive the depth position without ambiguity [11, 36].

For the particle image detection and evaluation first standard image processing tools like background subtraction and segmentation by a global or local threshold were used to determine regions of possible particle images. For the accurate determination of the x and y position of the particle images and respective diameters a_x and a_y, we will used here two methods, denoted as "MF Gauss" and "MF Countour", respectively.

The first method (MF Gauss) uses a cross-correlation with a Gaussian template, is very fast, and was shown to be very accurate if the model function holds and the intensity distribution of the particle images can be approximated by Gaussian functions [8]. However, this approach can show large errors in cases where the particle images show significant deviations from the Gaussian intensity

distribution due to diffraction effects or optical aberrations [20, 23, 37]. The second method (MF Contour) detects the edges of the particle images using the locally normalized intensity values at the rim of the segmented zone. This approach works especially well when the particle images are large and show a distinct edge or bright outer ring. For details the interested reader is referred to Fuchs et al (2016) [13, 14].

Using the width a_x and height a_y of the particle images, a parametric fit function based on Eq. (1) can be used to determine the z position [36]. More specifically, the fit function is here a curve representing a_x and a_y as a function of z. Given a measured particle image with diameters $[a_x, a_y]$, the corresponding z is obtained as the value that minimizes the Euclidean distance between the measured diameters and the fit function. This distance can be used as a validation criterion, thus rejecting particles that have diameters too far from the fit function [9]. In real experiments, as the case discussed in Section IV B, the fit function often differs across the image sensor due to optical aberrations, therefore a mapping of different fit functions across the xy-plane is performed to account for that.

B. Normalized cross-correlation with calibration image templates (CC)

Since its simple presentation in 1995 by Lewis [22], the normalized cross-correlation has been a standard image processing tool for feature tracking and template matching such as in motion-tracking, facial recognition, and medical imaging [24, 38]. And since the introduction of defocusing-based particle tracking through direct comparison to calibration images, the normalized cross-correlation has been a common choice to rate the similarities between measured and calibration images, i.e. by using its maximum peak value C_m as the similarity coefficient ranging from 0 to 1 (1 for perfect match) [1]. Here, the use of a single similarity parameter C_m offers a simple and robust approach for the evaluation and rejection of outliers. Another key advantage of using the normalized cross-correlation in defocus particle tracking is that it has a low sensitivity to light-intensity fluctuations and variations which are common in defocusing-based particle tracking applications. However, due to its simplicity, it does not work well for complex image matching problems, i.e. if the image objects in the matched images are not similar. This could for example be due to the objects having different orientation or size, such as within the tracking of biological cells. Furthermore, the performing of normalized cross-correlations has high computational costs and it is therefore crucial to limit the number of cross-correlations e.g. using predictive correlation schemes [34].

We applied the cross-correlation-based approach (CC) using *DefocusTracker*, which is a freely-available and open-source Matlab implementation [2]. In particular we used "method_1" described in Rossi and Barnkob [34]. The datasets were evaluated for C_m values of 0.5 and 0.9 and by the use of sub-image interpolation. To reduce the effect of noise, an in-plane median filter of 5×5 pixel was applied. *DefocusTracker* allows for a higher detection-rate of particles with overlapping images by use of an iterative procedure [34]. However, such iterative procedure has not been implemented in the used MF and NN algorithms and will therefore not be applied in this work.

C. Deep neural networks (NN)

Deep neural networks (NN) have shown to be very successfully for recognizing features in images. We apply a cascade of two deep neural models to first detect particles, i.e. their xy-position, and to infer their depth, i.e. their z-position, from a 2D image. The first network is supposed to determine a bounding box for each individual particle in an image, thus identifying their xy-position as shown for instance by König *et al.* [21]. We used the Faster R-CNN object detection method [32] with a ResNet-101 feature extraction network for this purpose. During training, batch normalization after the convolutional layers of the feature extractor was applied to decouple Faster R-CNN's individual networks typically allowing a faster optimization. We optimized the networks with stochastic gradient descent (SGD) and a momentum of 0.9. The initial learning rate was set to $3e - 4$, which was further reduced by a factor of 10 after 900 thousand and after 1.2 million iterations.

For defocus and astigmatism particle tracking, their 2D features represent also their depth, thus, allowing a network to learn distinguishing particles of different heights. Predicting a continuous output value, such as the desired z-position per particle, based on a given input (particle image) constitutes a regression problem from a machine learning perspective. We used a second deep convolutional neural network (CNN) to obtain a dense feature representation of a particle on which we trained the relationship between the appearance of a particle and its z-position in the penultimate layers of the network. This model directly succeeds the previous object detection model, i.e. the first model detects individual particles and derives their xy-position that is then used to extract individual particle images as input to the second model.

We studied four different popular and up-to-date CNN topologies for this second depth regression network, i.e., VGG16 [40], VGG19 [40], EfficientNet-B0 [42], and NASNet-mobile [45]. We chose these comparably less complex and not so deep topologies since the given visual concepts to extract from particle images are considered rather basic. We trained each network with an initial learning rate of 0.01, which reduced by a factor of 10 as soon as the validation loss does not improve for five consecutive epochs. In total, we trained each model for 200 epochs with a batch size of 16 images and applied

the Adam [19] optimizer. Later, we used early stopping to select the model with the lowest validation loss and reported all inference results based on this model. We teached networks with a mean square error (MSE) objective function since we were tackling a regression problem and used the mean absolute error (MAE) metric for measuring a networks' performance on test datasets.

TABLE I. Performance, training time and trained epochs across the three astigmatism degrees and four CNN typologies

Astigmatism degree	Topology	Training		MAE
		epochs	time [min]	
no	VGG16	16	120	**0.0110**
	VGG19	18	120	0.0141
	EfficientNet-B0	25	240	0.0420
	NASNet-mobile	38	200	0.0430
mild	VGG16	15	120	0.0050
	VGG19	16	120	**0.0047**
	EfficientNet-B0	26	240	0.1210
	NASNet-mobile	35	180	0.0225
strong	VGG16	15	120	0.0152
	VGG19	17	120	**0.0106**
	EfficientNet-B0	30	240	0.0700
	NASNet-mobile	37	200	0.0197

The neural network were trained on 17,328 labelled synthetic particle images, taken from the datasets with no noise and no overlapping particles (see Section III A). The results across the different trained models are reported in Table I. From left to right, the columns refer to the degree of astigmatism in the dataset, the network's topology, epoch that yielded the lowest validation error, the training time up to this epoch, and the MAE across all test samples. We found that the VGG topologies yield the best performing model. Thereby, the no astigmatism dataset contains less visual concepts making the more shallow VGG16 the better choice, while the astigmatism datasets benefit from the deeper VGG19 topology with its extra layers of convolution operations. Even the smallest of the EfficientNet topologies B0 with its 18 layers as well as the NASNet-mobile, both the result of a systematic topology optimization, seem to have prohibitive much training capacity for the given task resulting in considerably worse performance than the VGG topology.

For the results in Section III C, we trained three object detection models and applied Faster R-CNN as described above. These models were trained with 384 individual particle images of the respective astigmatism degree coming from all noise levels and using the datasets with no particle overlap and with $N_S = 0.1$ (200 particles per image). Additionally, we trained three depth regression models, one per degree of astigmatism with the respective best performing topology discussed above, i.e., VGG16 for no astigmatism and VGG19 for mild and strong astigmatism. These models were trained with 51,984 individual particle images of the respective astigmatism degree coming from all noise levels and using the datasets with no particle overlap.

To be successfully trained, NN methods require a large amount of labelled data which are easily accessible with synthetic images. However, in case of experimental images the labelled data (i.e. the true values of x, y, and z positions) must be obtained with a different measurement method. This makes the uncertainty estimation much more complex, therefore in this work we limit the analysis of NN methods only to the synthetic cases.

III. ANALYSIS OF SYNTHETIC IMAGES

The assessment of the methods is performed on synthetic images, following a well-established practice in particle tracking studies [15, 16, 31]. Synthetic images give access to the ground truth of the data (in our case the three-dimensional coordinates x, y, and z of the particles), therefore allowing for more objective comparison of different algorithms or methodologies. Furthermore, it is crucial when working with neural networks, which by definition require a large number of labelled data to be successfully trained. This requirement poses of course several challenges to the application of neural networks to real experiments, and we will discuss this aspect at a later stage in this paper. Instead, the objective of this section is to assess the performance of the different methods with respect to two major parameters [3]:

- The measurement uncertainty, σ_x, σ_y, and σ_z, for the determination of the in-plane coordinates (x, y) and for the out-of-plane coordinate (z), respectively. The measurement uncertainties are calculated on the entire depth of volume and are defined as the root-mean-square of the error between the measured coordinates and their ground truth. To investigate the measurement performance as a function of the depth coordinate z, we use the local depth coordinate uncertainty $\sigma_z^\delta(z)$ which gives the uncertainty on a number of bins (20 in this work) of width $2\delta = 0.05h$.

- The recall parameter ϕ, which consider the ratio between the number of valid measured particles (true positive) and total particles (true positive + false negative, i.e. not detected particles).

The concept of valid measured particles is here intended as the final output of a measurement, including eventual validation steps. It is important to keep in mind that the uncertainty and recall are connected to each other, since a lower uncertainty can be obtained at the cost of a more strict rejection criterion for valid particles, and vice-versa. Therefore a fair assessment of the performance must take into account both parameters, especially in the cases with many overlapping particle images.

For both MF methods, the validation criterion is based on the distance to the calibration curve (see Section II A and Ref. 9) which was set to 8 pixels in all cases. For the CC method, two thresholds were considered, $C_m =$

(a)

No astigmatism, high noise, no overlap
Mild astigmatism, low noise, $N_S = 0.1$
Strong astigmatism, no noise, $N_S = 0.6$

(b)

No astigmatism

Mild astigmatism

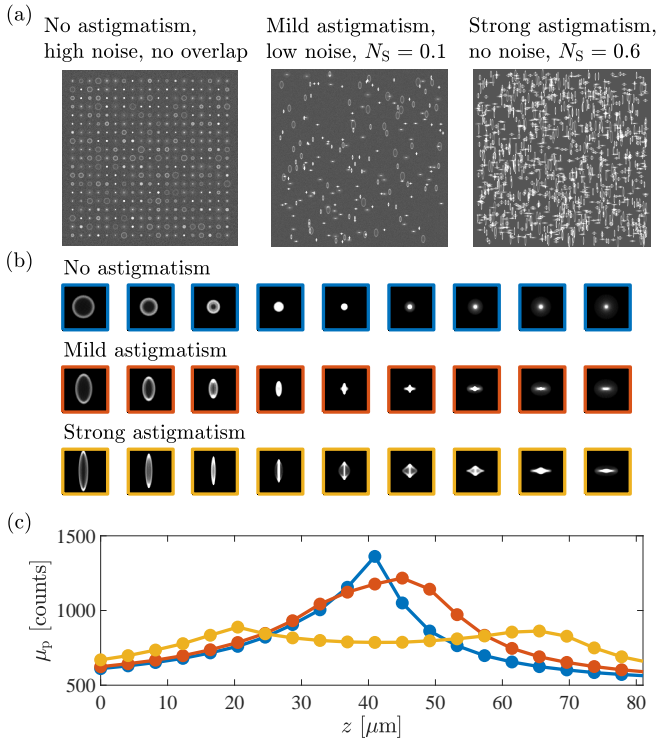Strong astigmatism

(c)

μ_p [counts] vs z [μm]

FIG. 1. Synthetic images: (a) Example field of view (FOV) of the synthetic images of different level of astigmatism, noise level, and particle image concentration N_S. (b) Example of the synthetic particle images at different z (columns) and for different level of astigmatism. (c) Mean particle image signal μ_p as a function of depth z.

0.5 (more valid particles, higher uncertainty) and $C_m = 0.9$ (less valid particles, lower uncertainty). For the NN methods there was no validation step and all the output data were considered valid measured particles.

A. Synthetic dataset preparation

The synthetic images are prepared using MicroSIG, a synthetic image generator for defocused and astigmatic particle images based on ray-tracing [33]. Following Barnkob and Rossi [3], we set the MicroSIG parameters to simulate 2-μm-diameter particles observed with a $10\times$ magnification objective showing spherical aberration, across a depth height of $h = 85$ μm. This setup is widely encountered in microfluidic experiments [8, 23] and has been selected as relevant benchmark case.

We consider three different optical arrangements, corresponding to three different degrees of astigmatism (none, mild, strong), shown in Fig. 1, in which the respective defocused particle images and their mean particle image signal μ_p are shown as a function of the depth coordinate z. The background level is fixed at 500 counts,

whereas the particle image intensity varies across the depth position ranging between 700 and 1400 counts.

For each image type, we consider three levels of noise (none, low, high). The low noise cases have a random Gaussian noise with $\sigma_I = 50$ counts. The high noise cases have $\sigma_I = 100 counts$, and additionally the image intensity of individual particle images fluctuate randomly between ± 50 % the nominal intensities. The corresponding signal-to-noise ratios (SNRs) are given by μ_p/σ_I [3].

Finally, we consider seven cases of different particle image density N_S, starting from a zero case where the particles are placed on a regular grid (no overlap), and gradually increasing the concentration up to a maximum of 1,200 particle images in $1,024\times1,024$-pixels images, corresponding to a N_S of approximately 0.6. The particle image density N_S, sometimes referred to as source density, is defined as the ratio between the sum of the particle image areas and full image area, for more details see Barnkob and Rossi [3].

Therefore, a total number of $3\times3\times7$ data-sets have been created. Each data-set contains a varying number of images in order to have an overall number of about 20,000 particle images, see Table II.

B. Uncertainty in the depth position for MF, CC, and NN

The synthetic-image datasets were processed with the three approaches with current state-of-the-art algorithms as described in Section II. The cases with no astigmatism were not processed with the MF methods, which rely on the presence of astigmatic aberrations. Throughout the analysis, we considered particles with depth coordinate uncertainty σ_z/h larger than 0.1 (so more than 10 % the full-depth scale) and in-plane uncertainty larger than 2 pixels as outliers, and count them as if they were undetected particles (i.e. false negatives and contributing to a lower recall ϕ).

We start out considering the error in the depth coordinate for the cases with no particle image overlap. Figure 2 shows the local depth coordinate uncertainty σ_z^δ as a function of z for the different levels of astigmatism and noise. In the following discussion, it is important to keep in mind that due to spherical aberration, particle images at lower depths show ring-shaped defocusing patterns, whereas particle images at larger z show Gaussian-shaped patterns [33]. As a consequence of this, the MF-Gauss method shows good performance in the middle-upper depth range where the particle images follow more closely the Gaussian intensity shape, but it has problems at lower depths or in the case of strong astigmatism where the images are more complex and show distinct local maxima of the intensity. The MF-Contour method in general performs very well with no noise, however it starts to have problems as long as noise is introduced, especially in the upper depths where the intensity gradients are smaller and it is more challenging to define

			Particles per image						
			361*	200	400	600	800	1000	1200
			Particle image density (ppp)						
			-	0.0002	0.0004	0.0006	0.0008	0.0010	0.0011
		⟨SNR⟩	Particle image density (N_S)						
No ast.	No noise	∞							
	Low noise	6.6	-	0.10	0.20	0.30	0.40	0.50	0.60
	High noise	3.3							
Mild ast.	No noise	∞							
	Low noise	7.7	-	0.09	0.19	0.28	0.37	0.47	0.56
	High noise	3.9							
Strong ast.	No noise	∞							
	Low noise	6.8	-	0.09	0.18	0.27	0.35	0.44	0.53
	High noise	3.3							

* Particle images are placed on a regular grid, no overlapping particle images are present.

TABLE II. Signal-to-noise ratio (SNR) and particle image density (N_S) for the $3 \times 3 \times 7$ synthetic image datasets.
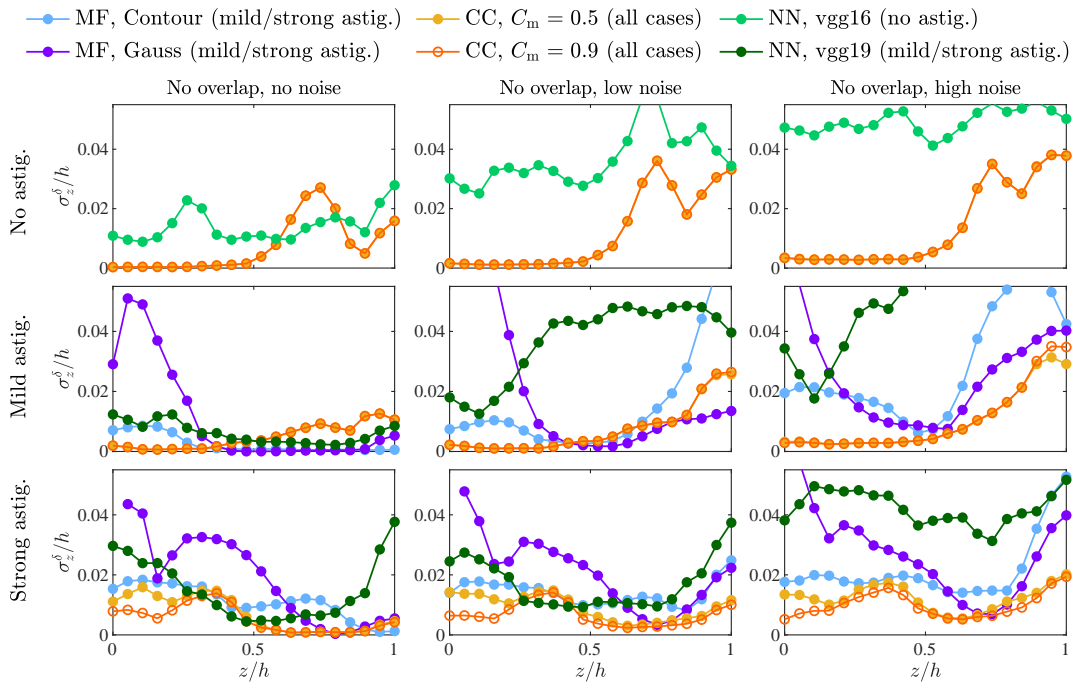

FIG. 2. Application of all three defocus tracking methods (MF, CC, NN) to synthetic-image datasets with no overlapping particle images: Local depth coordinate uncertainty $\sigma_z^\delta(z)/h$ as a function of the degree of astigmatism (rows) and noise level (columns). No-astigmatism cases were not processed with the MF methods as they rely on the presence of astigmatic aberrations.

the particle image borders.

The results of the CC method on the no-astigmatic case are consistent with previous observations [3], and show large uncertainties only in the focal region where it is difficult to differentiate the particle image shapes. This problem is mitigated by the introduction the astigmatic aberration, showing that even if an astigmatic aberration is not required for this methods, it is still advisable to avoid localized large errors in the focal region. The CC methods appears also to be robust against noise.

A more careful discussion must be done for the results presented for NN methods. The use of convolutional neural networks for defocus particle tracking is very promis-

ing, however it is still at its early stage. NN methods require a large amount of labelled data to be trained and this is experimentally challenging (experimental calibration images are typically available only for few hundreds discrete depth positions). Some strategies have been proposed to virtually increase the number of training data [12, 21], but we will not cover this aspect in the current comparison. We will instead consider a 'perfect set' of labelled data with no-noise (see Section II A) and train the network on them. The objective here is to evaluate how NN trained in optimal conditions compare with current state-of-the-art methods. The study here is also limited to the depth regression, and the xy-position of the parti-
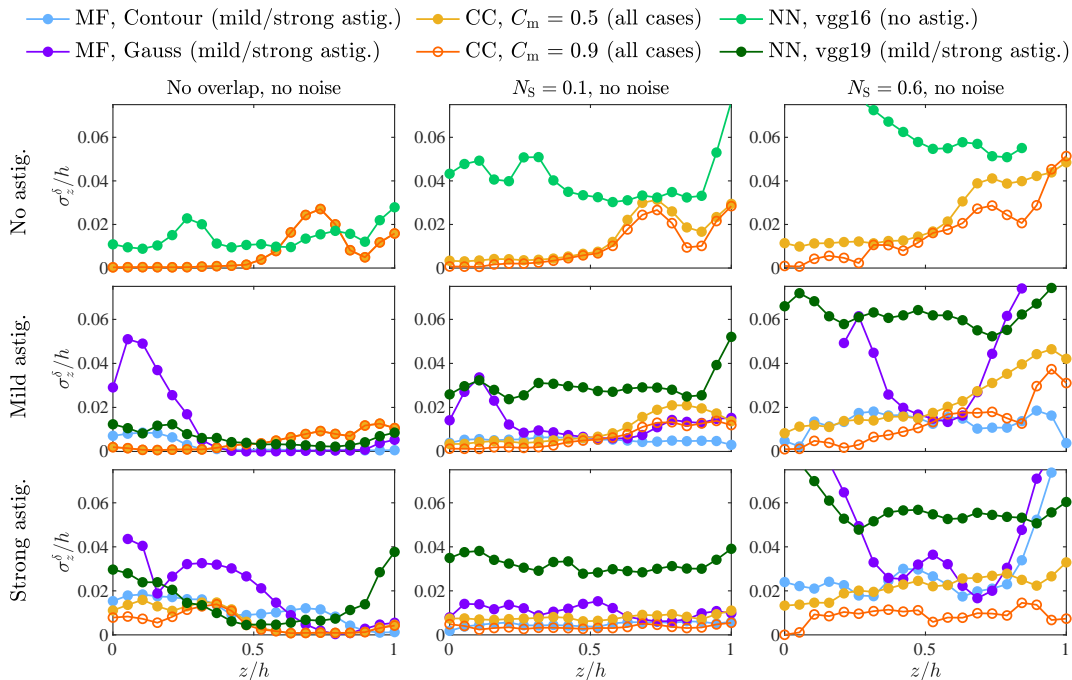
FIG. 3. Application of the all three defocus tracking methods (MF, CC, NN) to synthetic-image datasets with noise: Local depth coordinate uncertainty $\sigma_z^\delta(z)/h$ as a function of the degree of astigmatism (rows) and no overlap and particle image density N_S of 0.1 and 0.6 (columns). No-astigmatism cases were not processed with the MF methods as they rely on the presence of astigmatic aberrations. The missing points, for instance for MF Gauss, Mild astigmatism, $N_S = 0.6$, and $z/h < 0.25$, mean that no valid measured particles could be found in that region.

cle images is directly taken from the ground-truth data. We will show results including the in-plane identification and localization of particles in the next Section III C.

The results with no noise (first column in Figure 2), shows that NN methods are able to solve the task with a comparable accuracy of MF and CC methods, however the error is more evenly distributed across the depth. In agreement with the other methods, the setup with mild astigmatism is the one showing better performance. When noise is added, the uncertainty grows considerably, showing that a training in which also noise is present is probably necessary to improve the results.

Similar results are observed leaving out the noise but increasing the particle image density, as shown in Figure 3. Here, the difficulty is given by the increasing number of overlapping particles, which are clearly more difficult to process. Also in this case, CC methods appear to be the most robust, while the error increases significantly for NN methods. This is clearly expected since the training was performed on non-overlapping particles, so more sophisticated training schemes should be applied.

In conclusion, the results of this section show that MF methods perform very well for low concentrations and when the particle images fit well the used model function. CC methods are more robust for increasing noise and particle concentrations. NN methods are able to provide similar results than state-of-the-art methods, but it does not seem that they can improve significantly the

accuracy. This is a hint that the regression problem for the examined cases (no-noise, no-overlapping particles) is still relatively simple and NN methods do not offer significant advantages in comparison with conventional methods. A more interesting application of NN methods could be on cases with large particle image density or strong distortion across the image sensor, which are however beyond the scope of this work.

C. Overall uncertainty and recall for MF, CC, and NN

In the plots of Figs. 4, 5, and 6 we give a complete overview of the uncertainty and recall obtained for all the 54 datasets with overlapping particle images when using MF, CC, and NN methods. In the first column, the uncertainty of the out-of-plane component σ_z is reported. In the second column, the uncertainty of the in-plane components σ_x and σ_y is reported. In the third column, the recall parameter ϕ and the measured particle image density N_S', corresponding to ϕN_S, are reported. All plots are presented as a function of the particle image density N_S and parametrized for the 3 levels of noise.

The results confirm what observed in the previous section, with MF methods performing well at low noise and low particle image density (Fig. 4). The two MF methods perform similarly, but MF-Contour generally results
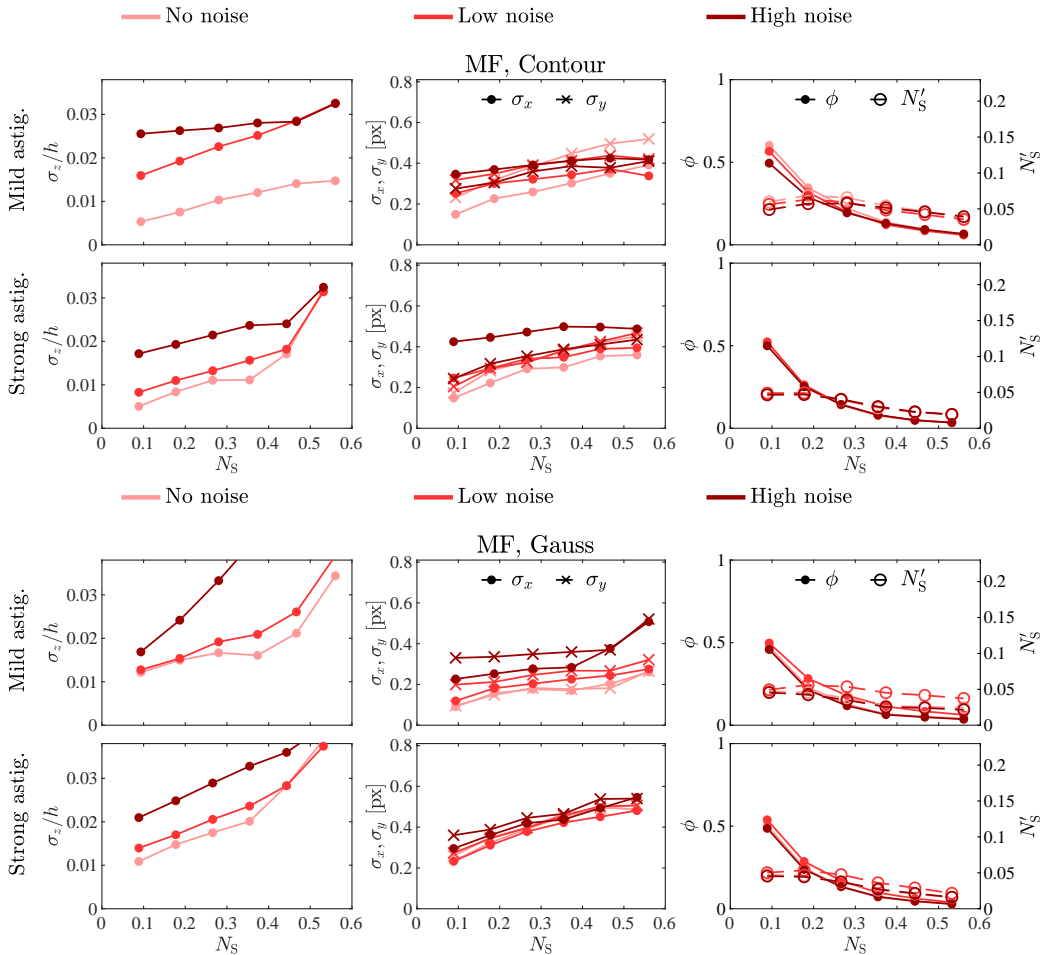
FIG. 4. Application of MF-based defocus tracking to synthetic-image datasets with mild and strong astigmatic aberration: Summary of depth coordinate uncertainty σ_z (first column), in-plane coordinate uncertainties σ_x and σ_y (second column), and recall ϕ as a function of particle image concentration N_S (third column). First and second row: MF-based approach with diameter detection based on contours. Third and fourth row: MF-based approach with diameter detection based on Gaussian template correlation.

in lower σ_z/h. The MF-Gauss approach is more robust against noise, but struggles for large particle image overlap.

For mild astigmatism and no noise, σ_z/h is below 0.005 (MF-Contour) and the in-plane uncertainty below 0.05 pixels (MF-Gauss). However, the uncertainty raises rapidly as the noise and particle image density are increased. The recall falls below 50 % already at $N_S = 0.1$. This is understandable, since those methods have not been designed or optimized to catch overlapping particles but to be used in low-seeding-density flows.

The minimal uncertainty for the CC results is also observed for mild astigmatism (no noise, low particle image density) with $\sigma_z/h = 0.06$ and in-plane uncertainty around 0.2 pixels (Fig. 5). The CC methods are however more robust against noise (almost no effect on the in-plane uncertainty) and are able to catch significantly more overlapping particles, with a recall still around 50 % at $N_S = 0.1$ for the case with $C_m = 0.5$. The uncer-

tainty can still be improved using a more strict validation criterion ($C_m = 0.9$), at the cost of a lower recall (but still larger than MF methods).

The 54 datasets were also processed using a NN method, considering this time an architecture of Resnet-101 that was able to identify particle images and their xy position and VGG16 and VGG19 that were able to identify particle depth z postition (Section II C and Ref. 21). Unfortunately, it was very problematic to train this architecture on these datasets and the results (reported in Fig. 6) showed much larger uncertainties and lower recall in comparison with MF and CC methods. We believe that the main problem here is the particle overlapping and the relatively small size of the particle images (around 50 pixels here against around 120 pixels in Ref. [21]). However, as already mentioned, NN methods are still on an early stage in defocus tracking and we believe that in the future different architectures and training strategies will be able to fill this gap.
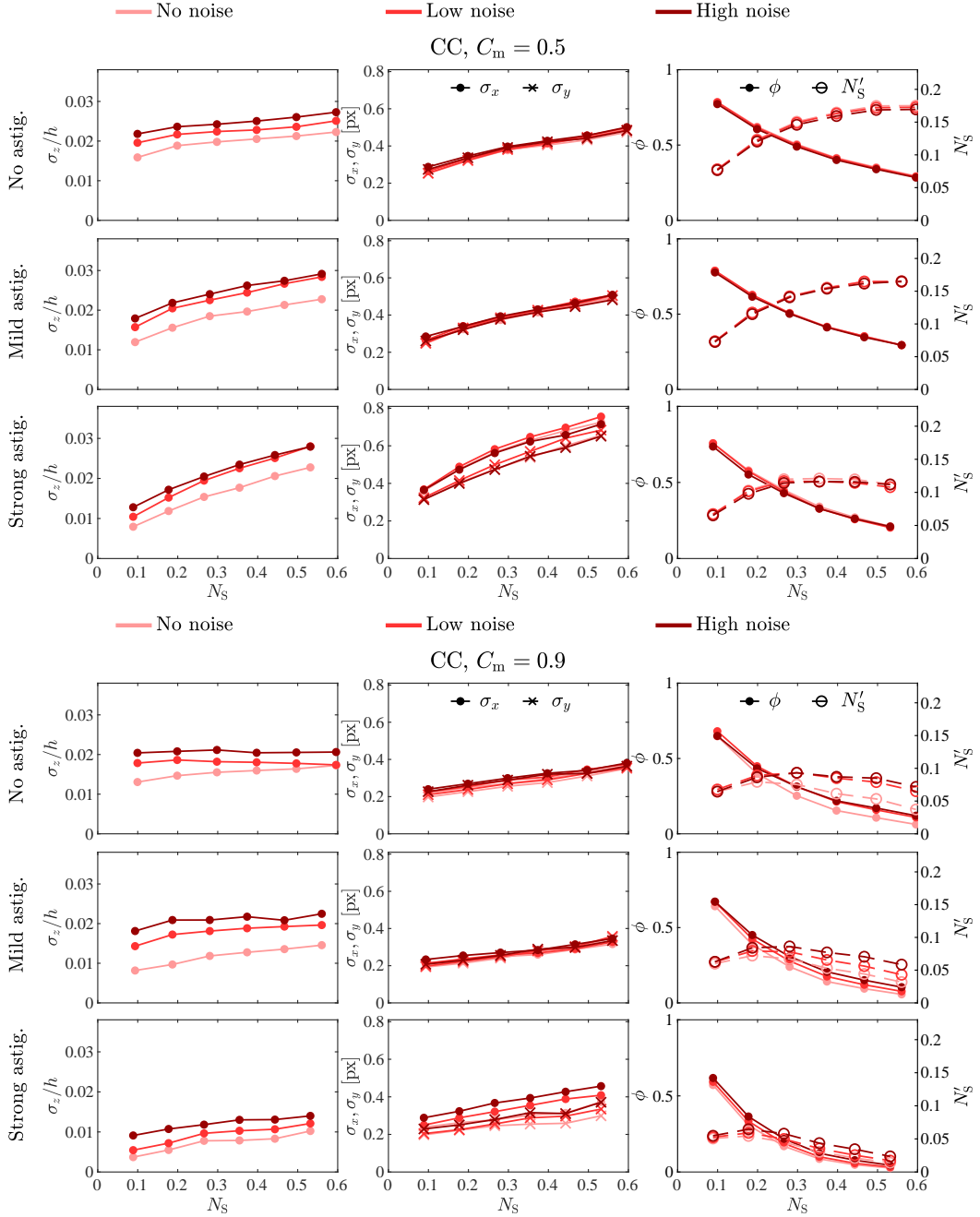
FIG. 5. Application of CC-based defocus tracking to synthetic-image datasets of different degree of astigmatic aberration: Summary of depth coordinate uncertainty σ_z (first column), in-plane coordinate uncertainties σ_x and σ_y (second column), and recall ϕ as a function of particle image concentration N_S (third column). The first, second, and third rows show the results with normalized cross-correlation peak parameter $C_m = 0.5$, while the third, fourth, and fifth rows show the results when using $C_m = 0.9$.

Finally, it is worth mentioning two points. First, as observed in the previous section, the optical arrangement showing the best results is the one with mild astigmatism. Interestingly, a similar conclusion based on a different analysis was obtained for classical APTV calibration methods [36]. Second, it should be noticed that the spatial resolution of a PTV analysis is given by the number of vectors per unit area, which is clearly directly proportional to the measured particle image density N_S'. Each configuration (method and dataset) has a maximum N_S', denoted as $N_S'^*$, corresponding to the critical particle image density N_S^* at which the method is able to catch more valid particles. Beyond that point the presence of overlapping particles will lower again the value of N_S'.
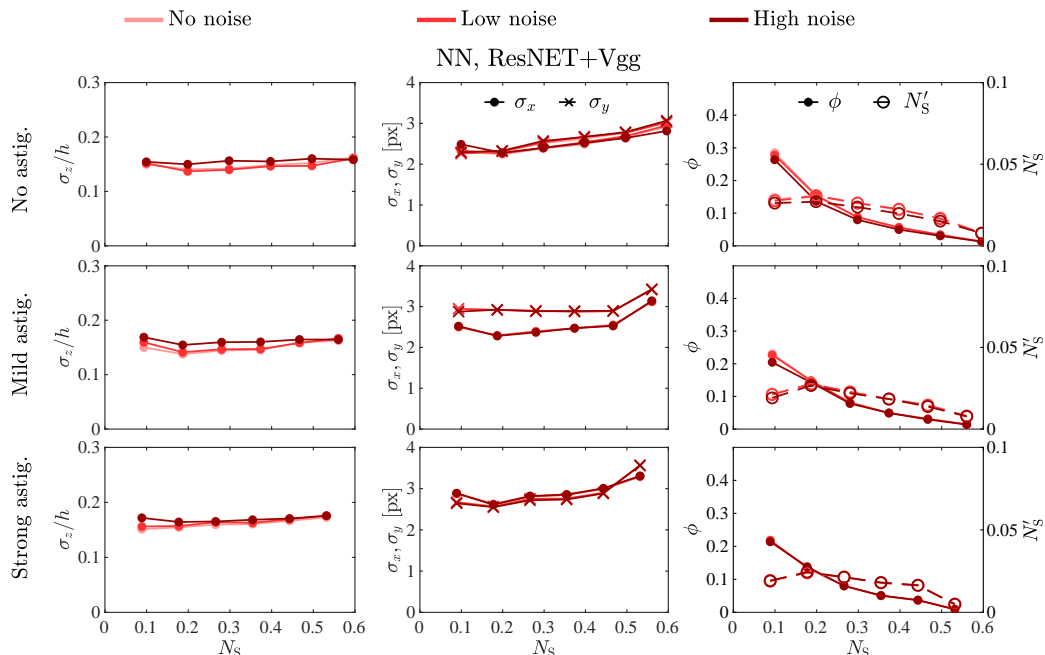
FIG. 6. Application of NN-based defocus tracking to synthetic-image datasets of different degree of astigmatic aberration: Summary of depth coordinate uncertainty σ_z (first column), in-plane coordinate uncertainties σ_x and σ_y (second column), and recall ϕ as a function of particle image concentration N_S (third column).
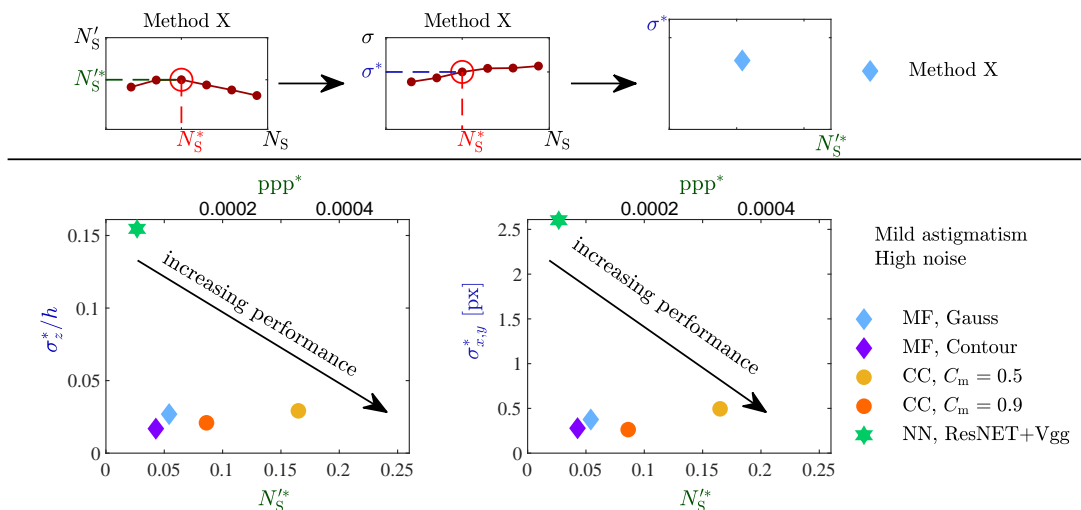

FIG. 7. Synthetic-image dataset challenge: Reference comparison of the different methods for the dataset with mild astigmatic level and high noise. The methods are compared with respect to uncertainties (depth and in-plane direction) versus the measured particle image density obtained at the critical particle image density N_S^* of each method, as shown in the sketch in the top. The dataset can be downloaded through https://defocustracking.com/.

Following these two points, in Fig. 7 we used the dataset of mild astigmatism and high noise for a concise comparison of the methods with respect to $N_S^{\prime*}$ (i.e. spatial resolution) and corresponding uncertainties σ_z^*/h and $\sigma_{x,y}^*$. The challenge here is to optimize the processing in terms of uncertainties (as low as possible) and critical measured particle image density (as high as possible). We propose this approach to establish an open quantitative

challenge to keep track of the improvements of defocus particle tracking methods. The current status sees CC methods giving the best performance in terms of $N_S^{\prime*}$, however at a cost of a larger uncertainty. The NN algorithms tested here performed worse in both metrics, showing that more research effort is needed to find more effective NN algorithms and training strategies. The dataset is available at https://defocustracking.com/.

IV. ANALYSIS ON EXPERIMENTAL IMAGES

A. Experimental dataset preparation

The experimental images were created using a similar setup as in König *et al.* [21]. For the calibration measurements particles were suspended in de-ionized water and filled in a sealed microfluidic chamber. As tracers fluorescent polysterol particles (530/607 nm, PS-FluoRed, MicroParticles GmbH) with a diameter of $2.5\,\mu$m were used. As the density of the particles is with 1.05 g/cm^3 slightly higher than the density of the water they will settle down on the microscopic slide. The water was maintained during the whole calibration experiment in the sealed microfluidic chamber. This avoids agglomeration, as typically appear when droplets containing particles were let to dry off. In addition the chamber ensures that the particle images are identical to later measurements and the particles do not move. This was important to quantify the position error in the xy-plane as will be outlined later. The assembled chamber was placed on top of an inverse microscope (Axio Observer 7, Zeiss GmbH) equipped with a long working distance Plan-Neofluar objective (M20x, NA=0.4, Zeiss GmbH) and a cylindrical lens of a focal length of $f_c = 250$ mm was placed in front of the camera at a distance of about 40 mm. A triggered cw-laser with a wavelength of about 532 nm (tarm laser technologies tlt GmbH & Co.KG) was used to illuminate the particles for 1 ms that were imaged on the sensor of an sCMOS camera (imager sCMOS, LaVision GmbH). The axis of the cylindrical lens was carefully aligned to the y-axis of the camera sensor. The different magnifications in the x- and y-direction were $50/130\,\mu$m/Px and $50/160\,\mu$m/Px, respectively. Particle images were captured in single-frame mode with a frame rate of up to 20 Hz. To increase the number of particle images for the different calibration methods a stage was used to shift the chamber in the xy-direction with a stepwidth of 70 μm from 0 to 700 μm, resulting in 121 images per xy-plane. In the z-direction the stages was moved over a distance of 80 μm using a stepwidth of 1 μm. This results in a total of $121 \times 81 = 9,801$ frames with approximately 35 particle image s on each frame. Exemplary particle images and the corresponding mean particle image signal can be seen in Fig. 8. Particle image overlap could not be totally avoided but was limited using this low particle image density of approximately 10^{-5} particles per pixel.

Note that in the processing of the experimental images, the $z = 0$ position is set arbitrarily to be that of the lowest scanned z-position while the remaining z-coordinates follow accordingly as we know the step size of the z scanning with high accuracy from the read-out of the motorized z-stage of the microscope.
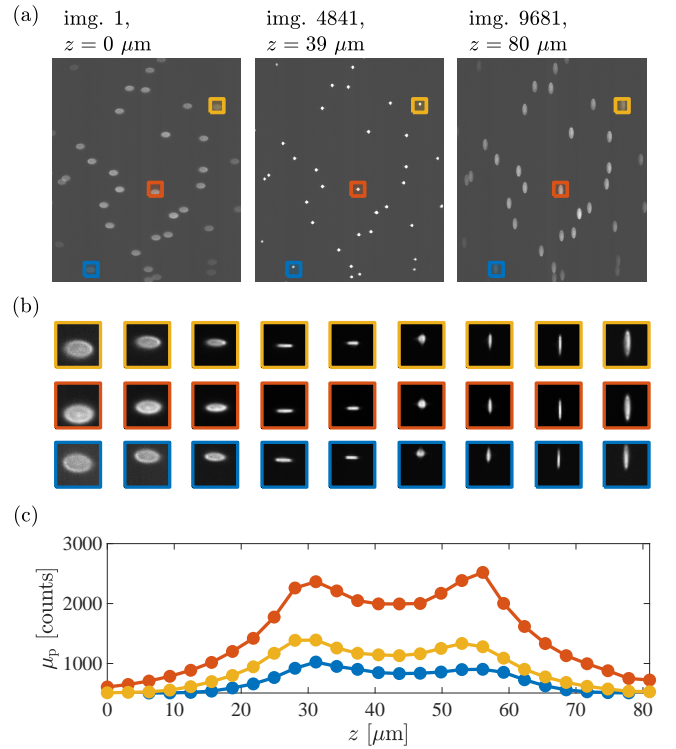

FIG. 8. Experimental images: (a) FOV at the same xy-position for z-positions corresponding to the lower bound (left), middle (middle) and upper bound (right) of the depth range. (b) Example of the experimental particle images at different z (columns) and xy positions (colored squares), respectively. (c) Mean particle image signal μ_p as a function of depth z.

B. Experimental-image processing results for MF and CC

The results obtained on experimental images are shown in Fig. 9. For the MF method, we used here the MF Gauss approach which provided the best results since the particle images has approximately a Gaussian distribution. A background subtraction was applied to remove an offset in the intensity level of the experimental images. For the CC method, we used the same settings as for the synthetic images, with a threshold of $C_\mathrm{m} = 0.95$. We did not used NN methods here since it was not possible to independently obtain the true values of the xy-coordinates for the training.

Additionally, it should be noted that experimental images present different shapes across the channel due to distortions and aberrations (see Fig. 8), therefore a correction method must be used to account for this. For the MF Gauss case, the images were divided into several square sub-regions with an edge length of 20 pixel for which separate calibration functions were determined to identify the z-position in correspondence to the xy-position. For the CC case, a calibration stack obtained from one particle image in the center of the calibration
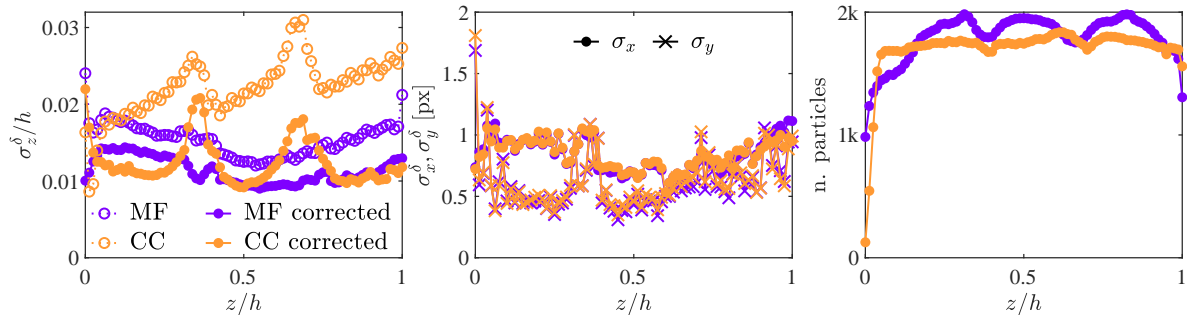
FIG. 9. Evaluation of experimental datasets using MF- and CC-based defocus tracking approaches: (left) Local depth coordinate uncertainty σ_z^δ, (center) local in-plane coordinate uncertainties $\sigma_x^\delta, \sigma_y^\delta$, and (right) local number of detected particles as function of the depth coordinate z/h.

images was used to calculate z-positions of different particles across the whole sensor in the calibration images (thus with known depth positions). A parabolic function was used to fit these data and later on to correct the measurements (more details on this procedure are given in [34]).

In the left part of Fig. 8 the error in the depth direction as a function of z/h is shown (both before and after correction). Both approaches result in reasonable mean errors over the z-position, which never exceed 3 % of the depth range, and after correction the mean error level is reduced to 0.0120 and 0.0114 for CC and MF, respectively. The error for the CC method is a little bit larger, especially in the region where the main focal planes are at $z \approx 30$ μm and $z \approx 65$ μm. In these regions the particle images shapes are very close among each other, therefore it is difficult for the cross-correlation to resolve the differences. On the other hand, in the MF method the correlation with a Gaussian function (MF) inherently smooths the signal and these peaks are less prominent and also the particle image deviations in the FOV become less important as can be seen on the lower error level of the non-corrected cases.

In the middle part of Fig. 9 the errors in the x- and y-directions are shown, respectively. As the ground truth of the x and y-positions was not known, these errors were obtained using the mean deviation from the displacement for calibration images taken when the stage with the calibration chamber was moved in the xy-plane. Since both approaches rely on determining the local maximum in the correlation plane with the particle image template for CC or the Gaussian template for MF, they both give very similar results and even the distribution of the errors fall almost on top of each other.

Since the ground truth of the number of possible particle images is not known in the experiment, on the right part of Fig. 9, the number of particles that are considered valid for each method are shown. Here CC shows a much more homogeneous flat distribution over the whole depth range, whereas for MF more valid particle images were detected in the central region. In total 136,387 and

158,496 particles are detected with CC and MF, respectively. The lower number of detected particles for the CC method is due to the large C_m threshold (0.95), and could be increased but at a cost of a larger uncertainty.

Finally, it should be noted that the present experimental images correspond to a case with low noise and almost no particle overlapping. This configuration is particularly favorable for the MF methods, as shown also in the synthetic cases. Provide an experimental case with overlapping particle images it is not straightforward and it is not presented here.

V. CONCLUSIONS

In this work we presented a comparison of three different methods for defocus particle tracking based on model functions (MF), cross-correlation (CC), and neural networks. Specifically, we evaluated the method performance based on the uncertainty and recall (i.e. the ratio between valid detected particles and total number of particles in the image). The analysis performed on synthetic and experimental images with different noise levels and particle image overlapping shows that:

- In general, all the three methods can be used for defocus particle tracking. However, the specific NN algorithms used in this work perform worse on all metrics in comparison with the more mature MF and CC approaches, showing that more work is needed to identify the more well-suited NN algorithms and training strategies.

- An astigmatic aberration is required only for MF methods, although it is advisable to introduce a mild astigmatism for all methods. This increases feature information in dependence on the depth position, which lowers the uncertainty close to the focal position. On the other hand, strong astigmatism often results in image aberrations that increase the uncertainty due to small local maxima in the intensity distribution.

- MF methods give the best performance in cases of low noise and low particle image overlapping. Using model functions like a Gaussian intensity distribution for the particle image detection works well, if the model represents the particle images well. In the case of ring-shaped particle images the error increases significantly as can be seen in Fig. 2. An advantage is, that this function works considerably well also without further image preprocessing, as the correlation or the fit by a Gaussian intensity distribution is inherently smoothing the results in the image plane.

- CC methods works very robust and reliable for all cases, and outperformed all other methods in cases with large noise and large particle image density. A major strength is that they work very robust determining reliable results even in heavily overlapping particle images. Inherently the cross-correlation amplifies the useful content in a signal and discards the rest.

- NN methods tested in this work generally performed worse in comparison with the MF and CC methods. However, NN methods are yet in an early stage within defocus tracking applications, whereas MF and CC algorithms and approaches are already in a mature phase. Given the huge success of applying NN methods in a wide variety of scientific fields, we do expect to see a future and rapid improvement of performance of NN methods in defocus tracking applications.

- Experimental images were tested with MF and CC methods on a case with no particle image overlapping. The results were in agreement with the synthetic cases, after suitable corrections were applied to account for distortion and aberrations across the field of view. The application of NN methods to experimental images requires strategies to provide large amount of labelled training data, which is currently an open open point and one of the future challenges in this field.

- Finally, as shown in Fig. 7, we propose a challenge for optimizing and comparing the processing performance of new and existing defocus tracking methods in terms of maximum measured particle image density (as high as possible) and corresponding uncertainties (as low as possible). The challenge uses a specific dataset (mild astigmatism, high noise) that can be downloaded through https://defocustracking.com/.

VI. AUTHOR CONTRIBUTION STATEMENTS

MR provided the synthetic test images, RB and MR performed the cross-correlation data analysis and prepared the figures, CC performed the model function analysis, MC and PM implemented the deep neural network and performed the image analysis. SS provided the experimental data and processed part of it. MR, RB and CC designed the concept, analyzed the data, and wrote the major part of the paper.

ACKNOWLEDGMENTS

[1] R. Barnkob, C. J. Kähler, and M. Rossi. General defocusing particle tracking. *Lab on a Chip*, 15(17):3556–3560, 2015.

[2] R. Barnkob and M. Rossi. Defocustracker: A modular toolbox for defocusing-based, single-camera, 3D particle tracking. *Journal of Open Research Software*, in review, 2020.

[3] R. Barnkob and M. Rossi. General defocusing particle tracking: fundamentals and uncertainty assessment. *Experiments in Fluids*, 61(4), 2020.

[4] S. Blahout, S. R. Reinecke, H. T. Kazerooni, H. Kruggel-Emden, and J. Hussong. On the 3D distribution and size fractionation of microparticles in a serpentine microchannel. *Microfluidics and Nanofluidics*, 24(3), mar 2020.

[5] P. Brockmann, H. T. Kazerooni, L. Brandt, and J. Hussong. Utilizing the ball lens effect for astigmatism particle tracking velocimetry. *Experiments in Fluids*, 61(2), feb 2020.

[6] Philipp Brockmann and Jeanette Hussong. On the calibration of astigmatism particle tracking velocimetry for suspensions of different volume fractions. *Experiments in Fluids*, 62(1):1–11, 2021.

[7] N. A. Buchmann, C. Cierpka, C. J. Kähler, and J. Soria. Ultra-high-speed 3D astigmatic particle tracking velocimetry: application to particle-laden supersonic impinging jets. *Experiments in Fluids*, 55, 2014.

[8] C. Cierpka and C. J. Kähler. Particle imaging techniques for volumetric three-component (3D3C) velocity measurements in microfluidics. *Journal of Visualization*, 15(1):1–31, 2011.

[9] C. Cierpka, M. Rossi, R. Segura, and C. J. Kähler. On the calibration of astigmatism particle tracking velocimetry for microflows. *Measurement Science and Technology*, 22(1):015401, 2010.

[10] C. Cierpka, M. Rossi, R. Segura, F. Mastrangelo, and C. J. Kähler. A comparative analysis of the uncertainty

of astigmatism-µPTV, stereo-µPIV, and µPIV. *Experiments in Fluids*, 52(3):605–615, 2011.

[11] C. Cierpka, R. Segura, R. Hain, and C. J. Kähler. A simple single camera 3C3D velocity measurement technique without errors due to depth of correlation and spatial averaging for microfluidics. *Measurement, Science & Technology*, 21:045401, 2010.

[12] S. Franchini and S. Krevor. Cut, overlap and locate: a deep learning approach for the 3D localization of particles in astigmatic optical setups. *Experiments in Fluids*, 61(140):140, 2020.

[13] T. Fuchs, R. Hain, and C. J. Kähler. In situ calibrated defocusing PTV for wall-bounded measurement volumes. *Measurement Science and Technology*, 27(8):084005, 2016.

[14] T. Fuchs, R. Hain, and C. J. Kähler. Uncertainty quantification of three-dimensional velocimetry techniques for small measurement depths. *Experiments in Fluids*, 57(5):73, apr 2016.

[15] C. J. Kähler, T. Astarita, P. Vlachos, J. Sakakibara, R. Hain, S. Discetti, R. La Foy, and C. Cierpka. Main results of the fourth International PIV Challenge. *Experiment in Fluids*, 57:97, 2016.

[16] C. J. Kähler, S. Scharnowski, and C. Cierpka. On the uncertainty of digital PIV and PTV near walls. *Experiments in Fluids*, 52(6):1641–1656, 2012.

[17] H Pin Kao and AS Verkman. Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position. *Biophysical journal*, 67(3):1291–1300, 1994.

[18] F. Kiebert, S. Wege, J. Massing, Jörg König, C. Cierpka, R. Weser, and H. Schmidt. 3D measurement and simulation of surface acoustic wave driven fluid motion: a comparison. *Lab on a Chip*, 17(12):2104–2114, 2017.

[19] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In Y. Bengio and Y. LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[20] A. Kloosterman, C. Poelma, and J. Westerweel. Flow rate estimation in large depth-of-field micro-PIV. *Experiments in Fluids*, 50(6):1587–1599, dec 2010.

[21] J. König, M. Chen, W. Rösing, D. Boho, P. Mäder, and C. Cierpka. On the use of a cascaded convolutional neural network for three-dimensional flow measurements using astigmatic PTV. *Measurement Science and Technology*, 31(7):074015, 2020.

[22] J. P. Lewis. Fast normalized cross-correlation. *Proceedings of Vision interface*, 10(1):120–123, 1995.

[23] R. Lindken, M. Rossi, S. Große, and J. Westerweel. Micro-particle image velocimetry (µPIV): Recent developments, applications, and guidelines. *Lab on a Chip*, 9(17):2551, 2009.

[24] J. Luo and E. E. Konofagou. A fast normalized cross-correlation calculation method for motion estimation. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 57(6):1347–1357, 2010.

[25] R. Luo and Y. F. Sun. Pattern matching for three-dimensional tracking of sub-micron fluorescent paerticles. *Meas Sci Technol*, 22:045402, 2011.

[26] A. Marin, R. Liepelt, M. Rossi, and Christian J. Kähler. Surfactant-driven flow transitions in evaporating droplets. *Soft Matter*, 12:1593–1600, 2016.

[27] A. Marin, M. Rossi, B. Rallabandi, C. Wang, S. Hilgenfeldt, and C. J. Kähler. Three-Dimensional Phenomena in Microbubble Acoustic Streaming. *Phys Rev Applied*, 3:041001, 2015.

[28] C. D. Meinhart and S. T. Wereley. The theory of diffraction-limited resolution in microparticle image velocimetry. *Measurement Science and Technology*, 14(7):1047–1053, jun 2003.

[29] M. G. Olsen and R. J. Adrian. Out-of-focus effects on particle image visibility and correlation in microscopic particle image velocimetry. *Experiments in Fluids*, 29(7):S166–S174, 2000.

[30] J. S. Park and K. D. Khim. Three-dimensional micro-PTV using deconvolution microscopy. *Exp Fluids*, 40:491–499, 2006.

[31] M. Raffel, C. E. Willert, F. Scarano, C. J. Kähler, S. T. Wereley, and J. Kompenhans. *Particle image velocimetry: a practical guide*. Springer, 2018.

[32] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015.

[33] M. Rossi. Synthetic image generator for defocusing and astigmatic PIV/PTV. *Measurement Science and Technology*, 31(1):017003, 2020.

[34] M. Rossi and R. Barnkob. A fast and robust algorithm for general defocusing particle tracking. *Measurement Science and Technology*, 32(1):014001, 2020.

[35] M. Rossi, C. Cierpka, R. Segura, and C. J. Kähler. Volumetric reconstruction of the 3D boundary of stream tubes with general topology using tracer particles. *Measurement Science and Technology*, 22(10):105405, aug 2011.

[36] M. Rossi and C. J. Kähler. Optimization of astigmatic particle tracking velocimeters. *Experiments in Fluids*, 55(9), 2014.

[37] M. Rossi, R. Segura, C. Cierpka, and C. J. Kähler. On the effect of particle image intensity and image preprocessing on the depth of correlation in micro-PIV. *Experiments in Fluids*, 52(4):1063–1075, sep 2011.

[38] J. N. Sarvaiya, S. Patnaik, and S. Bombaywala. Image registration by template matching using normalized cross-correlation. In *2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies*, pages 819–822. IEEE, 2009.

[39] R. Segura, M. Rossi, C. Cierpka, and C. J. Kähler. Simultaneous three-dimensional temperature and velocity field measurements using astigmatic imaging of non-encapsulated thermo-liquid crystal (TLC) particles. *Lab Chip*, 15:660–663, 2015.

[40] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[41] W. Stolz and J. Köhler. In-plane determination of 3D-velocity vectors using particle tracking anemometry (PTA). *Exp Fluids*, 17:105–109, 1994.

[42] M. Tan and Q. V. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946, 2019.

[43] T.Weier, D. Baczyzmalski, J. Massing, S. Landgraf, and C. Cierpka. The effect of a lorentz-force-driven rotating flow on the detachment of gas bubbles from the electrode surface. *International Journal of Hydrogen Energy*,

42(33):20923–20933, aug 2017.

[44] M. H. Winer, A. Ahmadi, and K. C. Cheung. Application of a three-dimensional (3D) particle tracking method to microfluidic particle focusing. *Lab Chip*, 14:1443–1451,

2014.

[45] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le. Learning transferable architectures for scalable image recognition. *CoRR*, abs/1707.07012, 2017.