Particle detection by means of neural networks and synthetic training data refinement in defocusing particle tracking velocimetry

To cite this article: Maximilian Dreisbach *et al* 2022 *Meas. Sci. Technol.* **33** 124001

OPEN ACCESS
IOP Publishing
Measurement Science and Technology
Meas. Sci. Technol. **33** (2022) 124001 (17pp)
https://doi.org/10.1088/1361-6501/ac8a09

Particle detection by means of neural networks and synthetic training data refinement in defocusing particle tracking velocimetry

Maximilian Dreisbach[1,*], Robin Leister[1], Matthias Probst[2], Pascal Friederich[3,4], Alexander Stroh[1] and Jochen Kriegseis[1]

[1] Institute of Fluids Mechanics (ISTM), Karlsruhe Institute of Technology (KIT), Kaiserstraße 10, 76131 Karlsruhe, Germany
[2] Institute of Thermal Turbomachinery (ITS), Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, 76131 Karlsruhe, Germany
[3] Institute of Theoretical Informatics (ITI), Karlsruhe Institute of Technology (KIT), Am Fasanengarten 5, 76131 Karlsruhe, Germany
[4] Institute of Nanotechnology (INT), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

E-mail: maximilian.dreisbach@kit.edu

Received 31 May 2022, revised 2 August 2022
Accepted for publication 16 August 2022
Published 8 September 2022

Abstract

The presented work addresses the problem of particle detection with neural networks (NNs) in defocusing particle tracking velocimetry. A novel approach based on synthetic training data refinement is introduced, with the scope of revising the well documented performance gap of synthetically trained NNs, applied to experimental recordings. In particular, synthetic particle image (PI) data is enriched with image features from the experimental recordings by means of deep learning through an unsupervised image-to-image translation. It is demonstrated that this refined synthetic training data enables the neural-network-based particle detection for a simultaneous increase in detection rate and reduction in the rate of false positives, beyond the capability of conventional detection algorithms. The potential for an increased accuracy in particle detection is revealed with NNs that utilise small scale image features, which further underlines the importance of representative training data. In addition, it is demonstrated that NNs are able to resolve overlapping PIs with a higher reliability and accuracy in comparison to conventional algorithms, suggesting the possibility of an increased seeding density in real experiments. A further finding is the robustness of NNs to inhomogeneous background illumination and aberration of the images, which opens up defocusing PTV for a wider range of possible applications. The successful application of synthetic training-data refinement advances the neural-network-based particle detection towards real world applicability and suggests the potential of a further performance gain from more suitable training data.

* Author to whom any correspondence should be addressed.

Keywords: defocusing particle tracking velocimetry, post-processing, particle detection, synthetic training data refinement, deep learning, neural network

(Some figures may appear in colour only in the online journal)

1. Introduction

A broad variety of particle imaging techniques exists in the fields of science and engineering. After the rise of digital imaging, particle image velocimetry (PIV) [1] has become a very beneficial and cost-effective method to acquire accurate velocity information with reasonable uncertainty margins, due to the possibility of correlation computation not only for one particle but for a particle ensemble. Given the Lagrangian frame of reference from the observation of single particles, particle tracking velocimetry (PTV) can provide a better understanding of physical phenomena involved in mixing processes [2] or e.g. the movement of bacteria [3]. For the measurement of three-dimensional flow topologies of small spatial scales similar to the light sheet thickness or for flows with large velocity gradients, the resolution of the standard planar particle imaging setup is not sufficient, since velocities are averaged over the light sheet thickness. Therefore, volumetric particle tracking techniques were developed, which either employ multiple cameras, for example 3D-PTV [4, 5] or tomographic PTV [6], or make use of defocusing, such as astigmatism PTV (APTV) [7] or defocusing PTV (DPTV) introduced by Willert and Gharib [8], as macroscopic defocusing approach. Note that the latter used a three-pinhole mask as aperture to generate three particle images (PIs) of the same out-of-focus particle on the camera sensor. The technique was adapted by research groups specialized in micro applications like Pereira *et al* [9]. Using the complete defocused image as estimate of the out-of-plane position was introduced by Wu *et al* [10]. The single-camera setup required for the defocusing approach allows for a simpler calibration and might even be the only viable option for measurement volumes (MVs) that are partially obstructed for all but one viewing angle. APTV requires an additional cylindrical lens, while DPTV can be realised through a standard planar PIV setup, however the evaluation of the experiments is not trivial.

The particles in the MV can be assumed to be point light sources, that are emitting rays of light through a spherical lens onto the camera sensor [10]. In context of DPTV, the defocused particles, therefore, appear as rings in the image plane, for which the diameter of the ring encodes the distance of the particle from the focal plane. Assuming geometrical optics and diffraction as Gaussian functions, the correlation of the PI diameter $d_\mathrm{e}(z)$ to the depth coordinate z can be described according to

$$d_\mathrm{e}(z)^2 = M^2 d_\mathrm{p}^2 + 5.95(M+1)^2 \lambda_e^2 f_\#^2 + \frac{M^2 z^2 D_\mathrm{a}^2}{(s_\mathrm{o}+z)^2}, \quad (1)$$

as elaborated by Olsen and Adrian [11].

The first term of equation (1) describes the magnification of the PI in geometrical optics as function of magnification M and particle diameter d_p. The second term represents the effects of diffraction, defined by the wavelength λ_e of emitted light from the particle and the focal number $f_\#$ of the objective lens, while the third term describes the growth of the particle-image diameter with an increasing distance z from the focal plane in geometrical optics, taking into account the lens entrance aperture diameter D_a and the object distance s_o. For a particular optical equipment the magnification M, aperture D_a and focal number $f_\#$ are constants and for microscopic optics in general the object distance s_o relative to the distance z of the particles to the focal plane is very large, i.e. ($s_\mathrm{o} \gg z$). For these conditions equation (1) simplifies to $d_\mathrm{e}(z) \propto \left(\mathrm{const.} + z^2\right)^{1/2}$ [11]. Furthermore for a sufficient distance to the focal plane the equation can be linearized, since the third term governs the equation. For optics with a high magnification spherical aberration influences the shape of the defocused PIs significantly [10], which is not considered in equation (1). Therefore, an individual calibration function has to be found for any particular optical equipment.

In order to determine the particle velocities, first the images are post-processed via a particle-detection step, in which the planar position and diameter are determined, followed by a matching of the PIs on two consecutive frames (see figure 1). Since its introduction by Willert and Gharib [8] in 1992, a multitude of different defocusing particle tracking methods was developed, which according to Barnkob *et al* [12] can be distinguished into methods based on model functions, cross-correlation methods and—more recently—neural network (NN) methods. Most methods of the former two approaches share a common strategy, in which, firstly, region proposals for particles are determined by an image segmentation on background subtracted images and secondly the location is refined with sub-pixel accuracy by a particle-detection algorithm. Generally, overlapping particles are avoided by a low seeding density, typically ranging from $n_{ppp} = 10^{-4}$ to 10^{-3} particles per pixel (ppp), which poses a limit to the resolution of the method. Furthermore most post-processing algorithms exclude overlapping PIs, as they often lead to erroneous measurements.

Model functions like the presented model of Adrian & Olsen (1) are often based on the assumption that the light intensity of the PIs has a Gaussian distribution [11, 13] and allow for the deduction of the particle-depth position from the measured particle-image diameter. Fuchs *et al* [14] obtain region proposals through an intensity threshold on average intensity filtered and binarized images. The three-dimensional position of the particle is then determined from the edges of

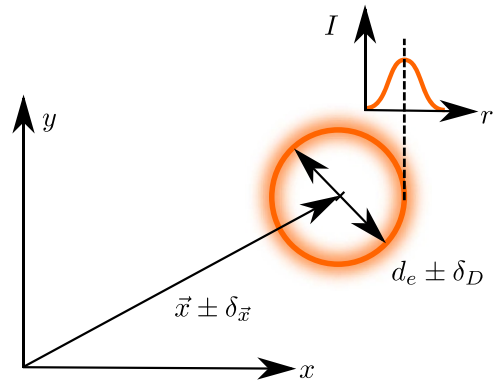


Figure 1. The planar position of the particle is defined by the centre-point location $\vec{x}$ of the defocused PI; the depth position is defined by the diameter d_e of the PI, measured at the maximum radial intensity.

the PI, which are detected by fitting the extrema of the PI in the horizontal and vertical direction with a thin-plate spline. Leister and Kriegseis [15] determine the three-dimensional position of the particle through a circular *Hough* transform [16] in one step. In a comparative study by Leister *et al* [17] this method—with an additional subpixel-refinement of the position and diameter and the edge detection method by Fuchs *et al* [14]—showed a comparable performance. Cierpka *et al* [18] determine the particle position by an auto-correlation of the proposed regions for particle candidates on unfiltered images. Barnkob *et al* [19] estimate the three-dimensional position of a particle by a normalised cross-correlation with a set of *a priori* determined calibration images at known depth location. In this approach sub-pixel accuracy is reached through a Gaussian fit of the in-plane pixel values and a parabolic fit of the cross-correlation peak along the calibration stack.

The established particle detection methods based on model functions suffer from a reduced accuracy for the cases where PIs diverge from the underlying limited theoretical assumptions, which describe the ideal physical principles involved in forming of the PIs. Methods based on cross-correlation however rely on PIs similar to the calibration templates, which limits their capabilities if the PIs vary significantly in planar direction. Adverse impacts from optical aberration, reflections, fluctuations in illumination and image noise lead to deviations in the PIs compared to the theoretical models or templates, which further reduces the accuracy and has a negative impact on the rate of successfully detected particles as well.

It has been shown that NN-based detection is more robust against the adverse impact of overlapping objects and low image quality [20] due to the ability of NNs to leverage a higher amount of optical features for the detection in comparison to the conventional algorithms. While image aberrations in general reduce the performance of conventional particle detection approaches, they represent features that the NN can use for the recognition of PIs and the subsequent determination of the depth position [12]. These features are internalised by the NN as defining characteristics of the PIs during machine learning on a set of labelled training images. Cierpka *et al* [21] show that a *Faster R-CNN* object detection algorithm [22], which is based on convolutional neural networks (CNNs) [23], can be successfully employed in the detection of particles on synthetic images in an APTV setup. Recently, König *et al* [24] found that a cascaded CNN developed on the basis of *Faster R-CNN* can be used to detect particles with high accuracy in APTV. Barnkob *et al* [12] use the *Faster R-CNN* object detection framework to determine the planar position of the particles and a subsequent CNN for the determination of the particle depth position from singular PIs, with CNNs trained on synthetic images. Franchini and Krevor [25] demonstrated an improved detection rate of overlapping PIs on images from APTV experiments with an object detector based on a long short term memory (LSTM) network combined with a CNN [26].

In contrast to the continuous development of NNs for APTV there has been a lack of application for DPTV so far. The accurate estimation of the depth position in DPTV relies only on the measurement of the PI diameter, whereas in APTV the orientation and size of the two axis of an elliptical PI can be exploited. Therefore, the strategies based on NNs for APTV have to be re-evaluated and adapted for particle detection in the DPTV context. In general the NN-based approaches require a large amount of labelled training data. High quality training data can in principle be acquired through manual image annotation. However, such an approach is unfeasible due to a high temporal effort. Alternatively, synthetic training data can be produced with a relatively low effort through model functions. Synthetic images, however, lack certain features of the PIs that are characteristic for the particular optical setup, which leads to a lower performance of the NNs once applied to real experimental datasets [27], because the training and test datasets stem from different distributions. Further improvement of the NN-based approaches for particle detection, accordingly, depends on the advances in the generation of suitable synthetic training data.

In view of the above-reported achievements and limitations, the objective of the present work is two-fold: Firstly and mainly the applicability of NNs for particle detection in DPTV is to be uncovered, complemented by an enhancement of the learning procedure to allow an advancement NN-based detection beyond synthetic test cases to a real world application. Therefore, a two-stage approach is introduced, in which first synthetic PIs are refined to match the characteristics of real image data by means of an image-to-image translation through a deep learning approach. Such refined datasets are hypothesised to serve as realistic training data, which enables the NNs for particle detection in the second stage to learn a more reliable detection of particles on real images from the experiment. As a second additional aspect, also the potential of NNs to overcome the currently open challenges in the post-processing of DPTV measurements on images with degraded quality or a high amount overlapping PIs is addressed, where the former likely occurs during experimentation and the latter is

particularly desired so as to achieve higher density of determined velocity information.

2. Methodology

The proposed two-stage approach is illustrated in figure 2 and bases on synthetic training data generation and refinement, as well as subsequent particle detection by means of deep learning on the recordings gathered by the DPTV experiments. An acquisition of training data through manual image annotation is impossible due to (i) the immense manual effort of labelling the required amount of images ($\mathcal{O}(10\,000)$) and (ii) the introduction of additional uncertainty and bias, which would reduce the detection accuracy for the NNs trained on those images. Therefore, the NNs for particle detection are trained on two kinds of synthetic and one semi-synthetic dataset for the extraction of particle position and size on real DPTV images [28]. The first synthetic dataset $\mathcal{D}_a$ is generated through conventional algorithms based on model functions, while the semi-synthetic dataset $\mathcal{D}_b$ is generated from cut-out real PIs annotated through a *Hough* algorithm. The refined synthetic dataset $\mathcal{D}_c$ is generated by means of an unsupervised image-to-image translation from the simple synthetic $\mathcal{D}_a$ and real $\mathcal{D}_b$ PI through a NN-based algorithm, as also indicated in figure 2.

2.1. Revolved 1D-Gaussian synthetic PIs

Since the theoretical considerations (such as [11, 13]) did not model the intensity distribution of a defocused PI, a different option to gain an initial synthetic PI could be chosen. One option would be the image generation with the help of a synthetic particle generator [29], the other to mimic already obtained PIs from an experimental set-up. The intention of this work was to model realistic-looking images with the use of a 1D-Gaussian function, that revolved around an axis. To further approximate the occurrence of spherical aberration, a linear function is superimposed on the radial intensity profile, which results in a higher intensity on the inner side of the particle compared to the outer side. The resulting synthetic intensity profiles for various PI diameters are plotted in the normalised intensity diagrams of figure 3, where synthetic ($\mathcal{D}_a$) and real ($\mathcal{D}_b$) PIs appear as yellow and red lines, respectively. The decrease in the peak intensity and broadening of the radial intensity profile towards more defocused particles was further approximated by a linear fit on the experimental data. Examples for the resulting synthetic PIs in comparison to their real counterparts of similar diameters are further displayed in figure 4.

2.2. Semi-synthetic data through real PIs

The dataset of real PIs $\mathcal{D}_b$ builds upon DPTV measurements in an open wet clutch by Leister and Kriegseis [15]. The recorded PIs are detected with the afore-mentioned *Hough* transform and subsequently extracted separately from the raw DPTV-images. Overlapping and irregular PIs were filtered out through manual inspection.

The unlabelled dataset $\mathcal{D}_b$ is employed for image-to-image translation directly, since only unpaired input ($\mathcal{D}_a$ and $\mathcal{D}_b$) is required by the NN for the refinement towards $\mathcal{D}_c$ as outlined in section 2.3; see also first stage in figure 2. Later on, the real PIs $\mathcal{D}_b$ have to be labelled as well for the training of the particle detection NNs (section 2.4, second stage in figure 2). In this case, these NNs will demonstrate a reduced spatial accuracy in comparison to the NNs trained on synthetic data, due to errors introduced by the labelling process.

However, NNs for particle detection trained on the semi-synthetic dataset $\mathcal{D}_b$ are expected to reach a high detection rate, since the feature distribution of $\mathcal{D}_b$ equals the distribution of the test data and accordingly will be used as a benchmark for the synthetic datasets $\mathcal{D}_a$ and $\mathcal{D}_c$. Consequently, the dataset $\mathcal{D}_b$ is only considered for the validation of the suitability of the two synthetic datasets as training data for the NN-based particle detection algorithms.

2.3. PI synthesis through unsupervised image-to-image translation

Particle detectors trained on purely synthetic PI-data are expected to have a reduced performance on the real experimental images, since they originate from significantly different data distributions. The reduced performance of NNs trained on synthetic data is a well documented and discussed observation [27]. To overcome this limitation, a new dataset is generated, which is foreseen to take combined advantage of the precise labels (location & diameter) from the synthetic training dataset $\mathcal{D}_a$ and the detailed feature content of the real PIs $\mathcal{D}_b$ from DPTV measurements. Particularly, a refined synthetic dataset $\mathcal{D}_c$ is developed by means of NN-based image-to-image translation between datasets $\mathcal{D}_a$ and $\mathcal{D}_b$, to provide fully labelled synthetic datasets with a more realistic image-feature distribution for the subsequent particle detection efforts (see section 2.4).

The available synthetic and real training data is inherently unpaired, because the particle datasets with a continuous range of diameters would need to be measured first in order to match them. This is not practical due to the introduction of measurement errors. As such, unsupervised methods for image-to-image translation are required to learn a mapping from an input domain consisting of the synthetic PIs to an output domain of real PIs.

Three state-of-the-art algorithms (*MUNIT* [30], *CycleGAN* [31] and *DRIT* [32]) for unsupervised image-to-image translation based on deep NNs are considered for the generation of refined synthetic PIs and evaluated in comparison to each other. All three methods are an extension of the generative adversarial network (GAN) concept [33], which has been shown to be very successful in generating synthetic images that closely resemble the real data they were trained on. *MUNIT* and *DRIT* additionally employ autoencoders [34], a type of NN that is commonly used for dimensionality reduction and data compression.

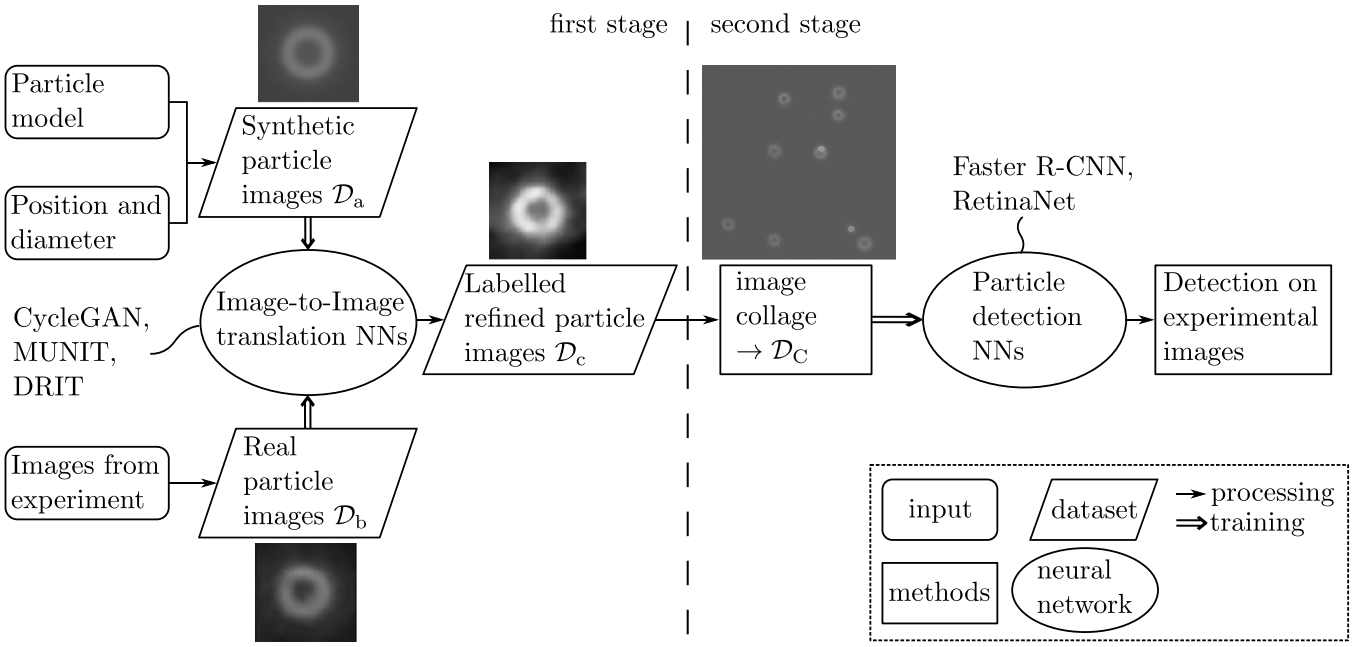


Figure 2. Two stage approach: integrated training pipeline for NN-based particle detection, employing image-to-image translation from unlabelled real and labelled synthetic data for synthetic training data refinement.

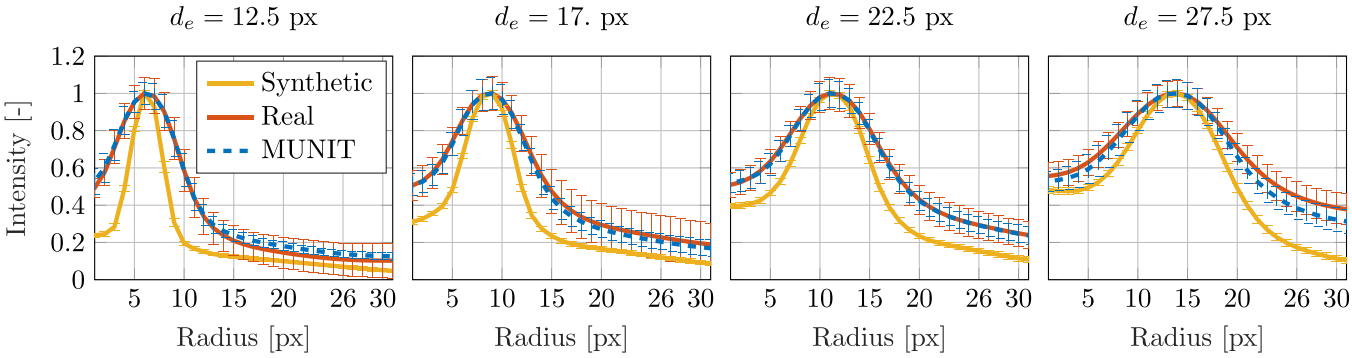


Figure 3. Comparison of the radial intensity profiles for synthetic $\mathcal{D}_a$ (■) and real PIs $\mathcal{D}_b$ (■), as well as PIs generated by *MUNIT* $\mathcal{D}_c$ (■, see discussion in section 2.3), for four characteristic diameters $d_e = 12.5, 17.5, 22.5, 27.5$ px. The intensity profiles are averaged over 100 PIs at each characteristic diameter $d_e \pm 0.025$ px and the intensity is normalised with the respective maximum intensities per diameter.

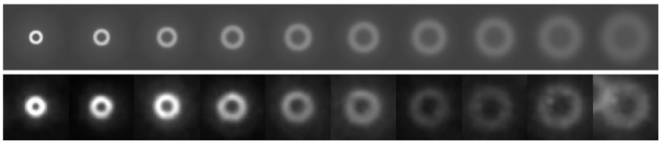


Figure 4. Comparison of randomly sampled synthetic $\mathcal{D}_a$ (upper row) and real $\mathcal{D}_b$ PIs (bottom row) at increasing diameters from $d_e = 10$ px to $d_e = 32.5$ px.

2.4. Particle detection by means of NNs

Generally, the scope of the particle detection step in PTV is the determination of the centre point location and the diameter of each PI on the recorded images. In the present approach two established standard methods for object detection based on CNNs—namely *Faster R-CNN* [22] and *RetinaNet* [35]—are considered for the particle-detection stage of the presented approach. Note that preliminary testings with the popular one-stage NN-based detection method *YOLOv3* [36] revealed a lower detection rate and accuracy than *Faster R-CNN* and *RetinaNet*, and hence was not further considered for the thorough comparison.

Previous reports by Cierpka *et al* [21] and König *et al* [24] demonstrate a good performance of the two-stage object detection algorithm *Faster R-CNN* for the post-processing of APTV experiments. *RetinaNet* belongs to the group of one-stage object detectors, representing a different approach to NN-based particle detection. Due to its feature-pyramid network based CNN architecture [37], *RetinaNet* is anticipated to reach a higher spatial accuracy than *Faster R-CNN*, as it has been shown to perform particularly well in detecting small objects [38]. Both *Faster R-CNN* and *RetinaNet* first process the input images through multiple convolutional layers to extract image features which are used for the detection. While this processing of the image features creates a so-called feature

hierarchy with increasing semantic information over the layers of the network, the resolution decreases in each consecutive layer. The detection is performed on the final layer, which contains the highest semantic information, but also the lowest resolution, which in turn leads to a reduction in spatial accuracy. The feature pyramid network (FPN) architecture encompassed in *RetinaNet* overcomes this down-sampling issue by establishing an inverse flow of information from later layers to earlier layers [37], which allows for the detection on higher resolution feature maps and offers the potential of a higher accuracy in comparison to the standard CNN-based detection approach.

The implementations of *Faster R-CNN* and *RetinaNet*, as well as the training algorithms were sourced from the *TensorFlow 1* library for Python [39], keeping default settings unless stated otherwise. Transfer learning based on NNs pretrained on the *COCO* dataset [40] was used to reduce the amount of required training data and accordingly training duration, and the risk of overfitting.

3. Data treatment and processing

In the present section, the proposed two-stage NN-based approach for particle detection in DPTV is first validated and tested with simple synthetic data. Subsequently, the impact of desired training-data refinement is evaluated in a comparative study on real data. The findings of each testing step will, therefore, initially be elaborated separately alongside the respective processing step and will be afterwards conflated in a more comprehensive discussion (see section 4). The performance of the proposed particle detection approach based on NNs is first assessed in section 3.1 on Gaussian synthetic images $\mathcal{D}_\mathrm{a}$ and compared to a conventional detection algorithm based on the *Hough* transform with a subsequent sub-pixel refinement of the particle position. An advantage of synthetic images is the possibility to associate the PIs with ground truth labels, which allows for an objective evaluation of the localization uncertainty by the different detection methods, as opposed to real images $\mathcal{D}_\mathrm{b}$ and refined synthetic images $\mathcal{D}_\mathrm{c}$. Recall that the labels for $\mathcal{D}_\mathrm{b}$ contain measurement errors and those for $\mathcal{D}_\mathrm{c}$ would compromise the labels by means of image translations.

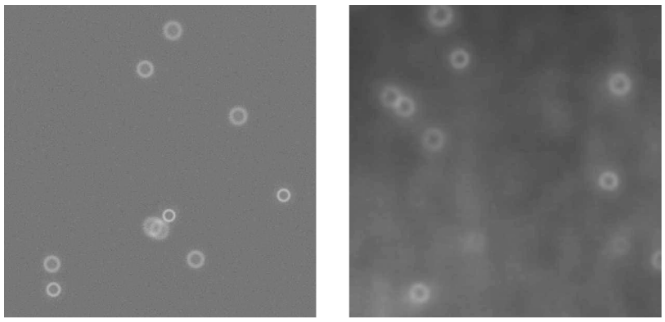
The ability of the NN-based approach for the detection on images with a high seeding density is evaluated in section 3.2, particularly on pairs of increasingly overlapping PIs. The NNs trained in section 3.1 are therefore employed on a test dataset with simple synthetic PIs $\mathcal{D}_\mathrm{a}$ with different relative overlap and the detection accuracy is evaluated. Likewise, the NN-based particle detectors trained in section 3.1 are tested under aggravated conditions in section 3.3, where the synthetic images are artificially superimposed with varying background-illumination conditions. Following the validation and testing of the NN-based detectors with simple synthetic data, the capability of the proposed approach based on synthetic training data refinement is evaluated. Therefore, first the results of the training data refinement by means of unsupervised image-to-image translation from synthetic PIs to the real domain are presented in section 3.4. The competing objectives of the transferability of features from $\mathcal{D}_\mathrm{b}$ to $\mathcal{D}_\mathrm{c}$ versus the simultaneous preservation of accurate labels from $\mathcal{D}_\mathrm{a}$ to $\mathcal{D}_\mathrm{c}$ are both evaluated here.

The generated refined synthetic data $\mathcal{D}_\mathrm{c}$ is then employed for the training of the particle detection NNs as presented in figure 2. Section 3.5, finally, addresses the detection performance of the NN-based particle detectors trained on refined synthetic data $\mathcal{D}_\mathrm{c}$ on real data. These detection results are compared to their corollary trained on simple synthetic data $\mathcal{D}_\mathrm{a}$, in order to determine the impact of the training data refinement. The comparison to particle detectors trained on cut-out real PIs $\mathcal{D}_\mathrm{b}$ additionally serves as a benchmark of the maximum possible detection performance with ideal training data, i.e. data that comes from the same distribution as the test set. Consequently, all three data sets $\mathcal{D}_\mathrm{a}$, $\mathcal{D}_\mathrm{b}$ and $\mathcal{D}_\mathrm{c}$ are considered to test the NN-based particle detectors *Faster R-CNN* and *RetinaNet* in a comparative manner on real data.

Unless stated otherwise the different synthetic image datasets for the particle detection study were created by placing ten PIs at random positions on a plain background, allowing for overlaps. The mean background intensity matched the background intensity observed in the experimental images [15] and the noise observed in the real images was approximated by white Gaussian noise with a signal-to-noise ratio (SNR) of 30. The images were saved as 8 bit .png-files with a resolution of 600×600 pixels. The NN-based detection approaches *Faster R-CNN* and *RetinaNet* were trained by supervised learning through stochastic gradient descent with momentum (SGDM) [41]. The NNs of all image-to-image translation models were trained with the Adam optimisation algorithm [42], which is an extension of SGDM. For completeness, the full set of training hyperparameters are listed in tables A1 and A2 in the appendix.

The performance of particle detection methods was evaluated by the metrics precision and recall, as well as the uncertainty and bias errors in localisation:

- The bias errors of the PI diameter δ_D and the planar position δ_X and δ_Y are calculated by the mean of the deviation $e_i = x_i - x_i'$ between the measured coordinates x_i' from the ground truth labels x_i in pixel (px) $\delta_i = \frac{1}{n}\sum_{i=1}^{n} e_i$ [43].
- The measured uncertainties σ_D, σ_X and σ_Y are calculated by the standard deviation of the errors between the measured coordinates and the ground truth labels in pixel (px) $\sigma_i = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - \delta_i)^2}$ [43].
- The mean absolute errors (MAEs) ε_D, ε_X and ε_Y are calculated by the mean absolute deviation of the measured coordinates from the ground truth label in pixel (px) $\varepsilon_i = \frac{1}{n}\sum_{i=1}^{n}|e_i|$ [44].
- The precision $P = \frac{TP}{TP+FP}$ describes the ratio of correct detections (true positives, TP) from all detections made by the particle detector on the test dataset. Therefore, it can be used as a measure for the ratio of false positive detections (FP) [45]. True positives are defined as detections with an intersection over union $IoU \geqslant 0.5$ [46], which describes the extent of overlap between the ground truth and detection bounding boxes.



(a) synthetic dataset (b) experimental dataset

Figure 5. Example image from the (a) synthetic dataset compared $\mathcal{D}_\mathrm{A}$ and (b) a recording from DPTV experiments [15]; contrast enhanced and brightened for better visibility.

- The recall $R = \frac{TP}{TP+FN}$ is defined as the ratio of true positive detections (TP) to the sum of true positives and false negatives (FN), i.e. the ratio of detected particles to all particles in the test dataset [45].
- The average precision $AP = \frac{1}{11}\sum_{r=1}^{11} \max_{\tilde{r}\geqslant r} P(\tilde{r})$ provides a combined measure for the object detection performance that summarises the relation of precision and recall by averaging the precision over eleven intervals of recall [46].

Recall and uncertainty have to be evaluated in respect to each other, as they are inversely correlated, depending on a rejection criterion. A stricter rejection criterion leads to higher accuracy at the cost of a lower rate of detected particles with the opposite effect for a less strict rejection criterion. The *Hough* transform algorithm uses a prescribed sensitivity parameter to determine valid detections, while the NN-based detection algorithms use a certainty score that is retroactively calculated for each detected particle by the NN as a rejection criterion. The detection performance of the NN-based detection algorithm is evaluated by plotting the precision and recall over the range of certainty thresholds in the detected particles, which accordingly can be performed *a posteriori*, while the sensitivity of the *Hough* algorithm has to be defined *a priori*.

3.1. Detection of synthetic PIs

The synthetic test dataset $\mathcal{D}_\mathrm{A}$ is composed by ten PIs with revolved 1D-Gaussian intensity distribution $\mathcal{D}_\mathrm{a}$, which were synthesised as described in section 2.1. The diameter of the synthetic PIs ranges from $d_\mathrm{e} = 18$ pixels to $d_\mathrm{e} = 35$ pixels and matches the diameter distribution observed in the experiments [15]. The generated images have a particle per pixel value of $n_\mathrm{ppp} \approx 2.8 \times 10^{-5}$, which results in 3.4% of overlapping PI pairs (as per definition in figure 8). An example image from the synthetic training dataset is shown figure 5 in direct comparison to a real image from DPTV experiments.

From the synthetic dataset $\mathcal{D}_\mathrm{A}$ $n_\mathrm{t} = 40\,000$ images were used for training the object detection networks and $n_\mathrm{e} = 1000$ images were reserved for the evaluation of the detection performance on unseen examples. The NN-based detection methods *Faster R-CNN* and *RetinaNet* were trained for 5 epochs and were afterwards evaluated in comparison to the model-based *Hough* transform on the test partition of the synthetic dataset $\mathcal{D}_\mathrm{A}$.

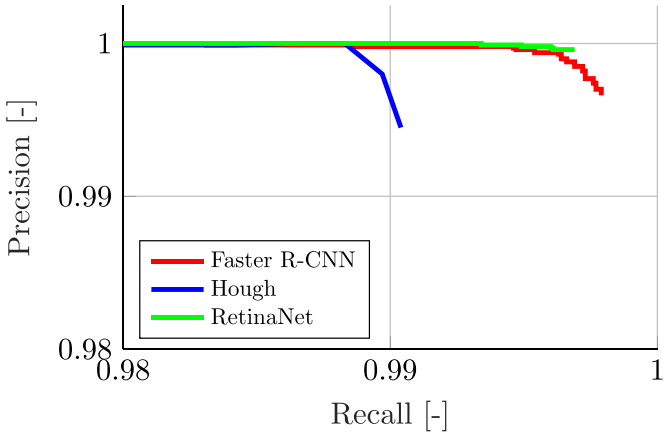


Figure 6. Comparison of the performance by the NN-based particle detection approaches *Faster R-CNN* and *RetinaNet*, and the *Hough* algorithm on Gaussian synthetic images $\mathcal{D}_\mathrm{A}$.

Table 1. Measurement errors of the NN-based particle detection methods trained and employed on synthetic DPTV data $\mathcal{D}_\mathrm{A}$ in comparison to the model based *Hough* algorithm.

IoU $\geqslant 0.5$	*Faster R-CNN*	*RetinaNet*	*Hough*
δ_D (px)	**0.158**	**0.158**	1.106
σ_D (px)	0.407	**0.296**	0.766
δ_X (px)	0.371	0.804	**0.063**
σ_X (px)	0.558	**0.253**	0.364
δ_Y (px)	0.129	0.064	**0.002**
σ_Y (px)	0.527	**0.228**	0.340

As can be seen in figure 6, the compared detection methods reached a high sensitivity on the synthetic test images, with a detection rate (recall) above 99% in combination with a low rate of false positive detections, expressed by precision values likewise above 0.99. Figure 6, therefore, might be generally considered a successful validation of the NN approach for DPTV particle detection. The detection performance of the NN-based approaches *Faster R-CNN* and *RetinaNet* is found to be minimally higher compared to the *Hough* transform, which mainly results from a better detection of overlapping PIs as will be further evaluated in the following section. The NN-based approaches only miss a third of the amount of particles relative to the *Hough* algorithm. While the effect is less relevant on simple synthetic images, it is expected to transfer and amplify on real images that are more difficult to detect for all methods.

As shown in table 1, the in-plane bias errors of *RetinaNet* δ_X and δ_Y significantly differ. This trend was observed over all tested datasets and for different random initialisations of the NN. Therefore the cause is suspected to stem from within the NN architecture. However for a velocity measurement in the context of particle tracking the bias error cancels itself and for the position measurement a calibration can be performed by an offset with the known bias error. Since such

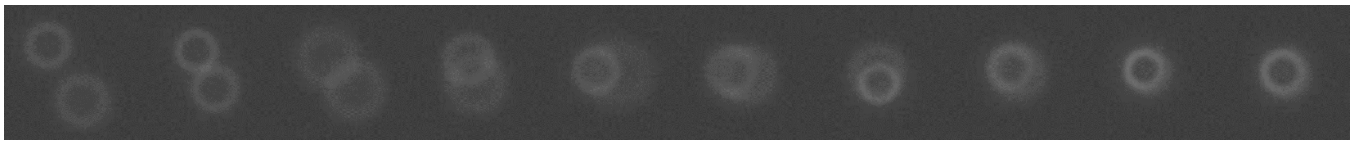


Figure 7. Synthetic Gaussian PI pairs at varying relative overlap.

(constant) bias does not influence the measurement accuracy of the particle tracking in any form, the following discussions only elaborate on the measurement uncertainty in terms of standard deviations. *RetinaNet* reaches the lowest uncertainty for the measurement of the planar position $\sigma_{X,Y}$ as well as the diameter σ_D of the PIs in comparison to *Faster R-CNN* and the *Hough* algorithm. Especially for diameter determinations *RetinaNet* shows a significant advantage, as obvious from the particularly small uncertainty compared to the *Hough* algorithm. This suggests in consequence, that significant improvements in the accuracy of the depth measurement can be realised by employing *RetinaNet* for particle detection. *Faster R-CNN*, as well, reaches a lower uncertainty in diameter determination, however combined with a higher planar localisation uncertainty compared to the *Hough* algorithm. In general synthetic images can reproduce arbitrary magnification and reproduction scales from experimental images. However, since the PIs from Leister and Kriegseis [15] are reproduced in the present work, an identical reproduction scale of 5.77 µm px^{-1} in planar direction and 35.3 µm px^{-1} in the depth direction, as well as a MV depth of 600 µm is consequently used to convert the uncertainties into the physical space as previously reported in [15]. Accordingly the NNs *Faster R-CNN* and *RetinaNet* have an average in-plane uncertainty $\sigma_{X,Y}$ of 3.1 and 1.4 µm, respectively and a depth uncertainty of $\sigma_D = 14.4$ and 10.4 µm, which corresponds to a relative uncertainty of $\sigma_D/h = 2.4\%$ and $\sigma_D/h = 1.7\%$, respectively. This compares to an in-plane uncertainty $\sigma_{X,Y}$ of 2.0 µm and a depth position uncertainty of $\sigma_D = 27.0$ µm or $\sigma_D/h = 4.5\%$ for the *Hough* algorithm. Overall the NN-based approaches are found to have a more balanced distribution of uncertainty between the determination of the planar and the out-of-plane position of the particles. In comparison to the *Hough* algorithm, this finding translates into an improvement in the accuracy of the depth coordinate measurement, which is generally an order of magnitude lower than planar accuracy with conventional approaches, thus potentially alleviating a downside of DPTV. The high detection rate in combination with a high spatial accuracy on synthetic PIs render the NN-based particle-detection approach a promising candidate for improvements in DPTV. In order to estimate the efficiency in a real-world setting, the detection performance on overlapping PIs and degraded images is evaluated in the following.

3.2. Analysis of overlapping particle-image pairs

A significant potential advantage of NN-based particle detection over conventional methods lies within the detection of overlapping PIs due to their ability to detect partly obscured

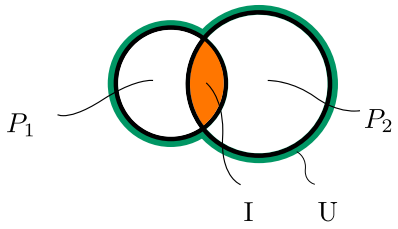


Figure 8. Illustration of the circular intersection over union (cIoU) for two PIs P_1 and P_2, which is used as the measure for the relative overlap of two PIs.

objects. Therefore, in the following section the capability of the NN-based approaches on overlapping PIs is investigated through a measurement of the detection rate and localisation accuracy for increasingly overlapping PI pairs and a comparison to the conventional *Hough* algorithm. For this purpose pairs of Gaussian synthetic PIs $\mathcal{D}_\mathrm{a}$ with increasing degrees of relative overlap were generated, as illustrated in figure 7. The relative overlap is measured by the circular intersection over union ($cIOU$), which is introduced as a specification of the commonly used intersection over union [46] for the context of circular PIs. The $cIoU$ is defined as the area of intersection I between the areas of two PIs P_1 and P_2, measured at the radius of peak intensity and divided by their union area U, i.e. $cIoU = \frac{P_1 \cap P_2}{P_1 \cup P_2} = \frac{I}{U}$, as also illustrated in figure 8. The test dataset consists of eight intervals of increasing cIoU, ranging from $cIoU = [0, 0.1]$, containing marginally overlapping particles to $cIoU = [0.7, 0.8]$, which included overlapping PIs that were not separable for a human observer, as can be seen in figure 7. Additionally, a test dataset of non-contacting PI pairs with relative distances between 5 and 0 pixels ($cIoU - c$) was included, since the results on the synthetic test dataset $\mathcal{D}_\mathrm{A}$ indicated an influence of such close pairs on the detection. Likewise, a set of test images containing PI pairs with larger relative distance >5 pixels ($cIoU - f$) was included to measure the base performance of the particle detection algorithms without overlap for comparison. Each dataset contains $n_\mathrm{e} = 1000$ test images, each of which is comprised of two randomly sized PIs within the considered diameter range of $d_\mathrm{e} = 18$–35 pixels.

The evolution of the recall for varying the overlap $cIoU$ is shown in figure 9(a) for the detection methods *RetinaNet*, *Faster R-CNN* and the *Hough* algorithm. As can be seen, both NN-based detection methods are able to detect and resolve both individual synthetic PIs in an overlapping pair at significantly higher degrees of overlap $cIoU$ than the *Hough* transform. In particular, the adverse impact of the overlapping

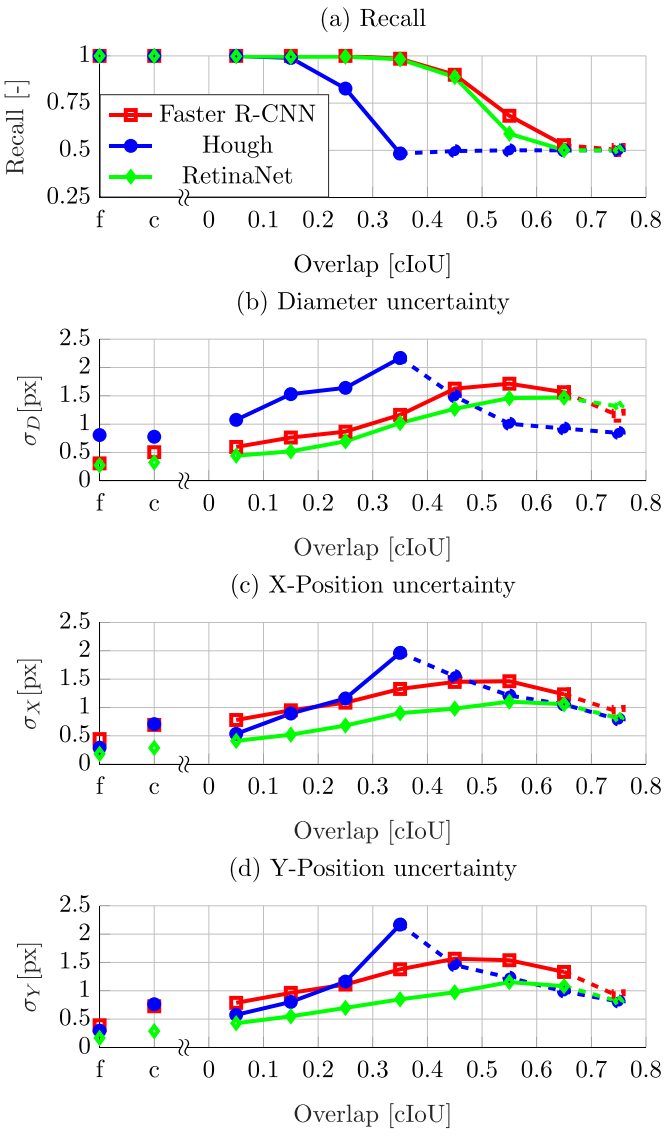


Figure 9. (a) Recall R; (b) uncertainty in diameter σ_D; (c) centre point X-position σ_X and (d) centre point Y-position σ_Y determination for PI pairs at increasing overlap measured by circular interception over union. The dashed lines indicate that only one PI was detected.

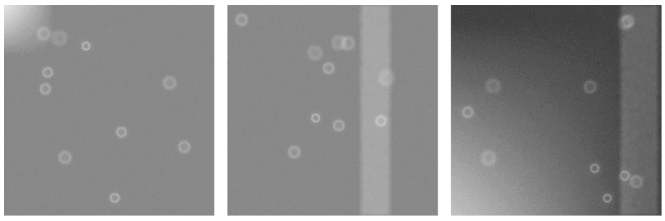
PIs on the detection can be neglected for *Faster R-CNN* and *RetinaNet* up to the range of $cIoU = 0.3$–0.4, as indicated by recall values above 0.98. The *Hough* algorithm, in contrast, already suffers from PI loss in the same *cIoU*-range, where on average only one of the two PIs among the respective pairs is identified—indicated by a recall of 0.5. It was further observed that the detection of overlapping PIs with similar diameters proved to be a more difficult task in comparison to pairs with a larger difference in diameter for all methods, resulting in both a lower detection accuracy and likewise a correspondingly reduced detection rate. A potential source of error for PIs with similar diameter could be that image features, such as intensity gradients of the two instances are similar and thus blend into each other, hindering a differentiation. The most common failure mode of the NN-based methods was that only one out of the two particles was detected, however with a relatively high spatial accuracy, while the *Hough* algorithm on the other hand in many cases detected a ghost particle in the intersection of the two particles, leading to additional large errors in the measured position.

Figures 9(b)–(d) shows the evolution of the measurement uncertainty for varying overlap. It should be noted that the evaluation of the uncertainty is only meaningful in the context of overlap for PI pairs with two detected PIs, i.e. recall of >0.5, and a maximum uncertainty is accordingly reached for this lower edge case of only one detected particle. For PI pairs with a higher degree of overlap the PIs merge more into each other and are therefore more likely detected as a single particle, thus reducing the measurement error for this detected particle. Consequently, the error estimation systematically looses the context of overlap, which is indicated by the dashed lines in figure 9. For all detection methods the uncertainty increases towards higher overlap and a negative impact from overlapping PIs is observed already at a lower degree of overlap in comparison to the discussed impact on the detection rate. This observation suggests a more sensible dependence of the localisation procedure on the overlap in comparison to the detection of the particle. As can be seen in figure 9(b) both NN-based approaches reach a significantly higher diameter-measurement accuracy on overlapping PIs in comparison to the *Hough* transform, with *Faster R-CNN* achieving half the uncertainty of the *Hough* algorithm in diameter determination and *RetinaNet* even slightly less. The planar accuracy of *RetinaNet* is considerably higher in comparison to the *Hough* algorithm, while *Faster R-CNN* reaches a comparable accuracy; see figures 9(c) and (d). The presented results demonstrate that NN-based particle detection allows for a significant increase in the accuracy of DPTV under the given conditions, especially for the out-of-plane measurement and overlapping particles images, therefore confirming the findings of section 3.1.

Furthermore it is shown, that a NN-based particle detection can resolve both PIs in an overlapping pair up to significantly higher overlaps than the conventional *Hough* algorithm, which suggests that an increased seeding density can be realised with the proposed approach. A larger amount of measurements from an increased seeding density would be especially advantageous for non-stationary flows and other experimental conditions in which ensemble averaging is not feasibly (if possible at all). In this context it is important to note that the present cIoU-based evaluation only addresses single events of overlap, which has to be considered for real experiments in combination with the respectively chosen seeding densities, as the latter determines the rate of overlapping PIs on the recorded images.

3.3. Detection on synthetic images with artificially degraded image quality

The adverse influence on the detection performance from an inhomogeneous background intensity over the DPTV images is evaluated on synthetic images with additionally superposed bands of higher intensity and large-scale intensity gradients



(a) gradient (b) stripe (c) both

Figure 10. Synthetic test images containing (a) an additional intensity gradient; (b) a stripe and (c) both adverse effects (contrast enhanced and brightened for better visibility).

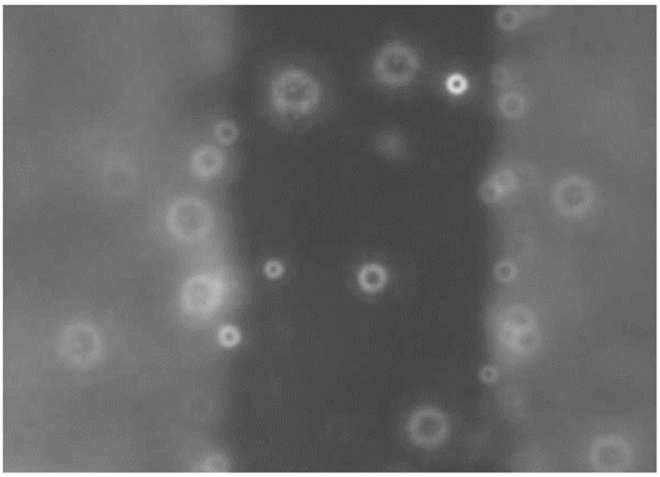


Figure 11. Recorded DPTV raw image above the grooved region of an open wet clutch.

as illustrated in figure 10. Changes in the background intensity resulting from the geometry of the MV, as observed in the experimental images e.g. figure 11 were simulated by randomly positioned vertical bands of higher intensity (figure 10(a)) and the effect of reflections as well as spatially uneven illumination over the MV (figure 10(b)) were simulated by intensity gradients, while in a third experiment both negative effects were superposed (figure 10(c)).

For each test case $n_e = 1000$ synthetic test images with each ten Gaussian PIs were produced. The performance of the detection methods on these three test datasets is compared for particle detectors trained on the synthetic DPTV dataset without a variation in the background intensity $\mathcal{D}_A$ in order to evaluate the behaviour of NN-based detection algorithms on degraded test data.

The robustness of all particle detection methods against image degradation is evaluated by means of precision-recall diagrams, as shown in figure 12. Obviously, both considered types of raw-image imperfections lead to reduced particle-detection performance, which becomes even more influential by the superposition of both negative effects. Interestingly, the behaviour of the two NN-based approaches was perceptibly different: While *Faster R-CNN* is found to be the most robust method against image degradation, with effectively unchanged precision and recall, *RetinaNet* produces both false positive as well as false negative detections in the regions of the image with perturbations of illumination homogeneity, as seen by the reduction in precision and recall.

For image regions, where an intensity band and gradient overlapped, the *RetinaNet* algorithm is unable to detect any particles, indicating that the NNs of *RetinaNet* react more sensible to differences between test and training images in comparison to *Faster R-CNN*. Synthetic image degradation leads to a reduced precision for the detection of the *Hough* algorithm, while the algorithms detection rate remains largely unaffected. The precision is reduced by the detection of ghost particles, mostly at the edges of the intensity bands that produce a sharp gradient in the background intensity on the image. Furthermore, regions of superimposed intensity gradients and stripes reveal an amplification of this effect.

For the *Hough* algorithm the majority of false positive detections can be avoided by means of an *a priori* adjustment of the sensitivity to a value of 0.7 from the initially chosen and quality-approved value of 0.9 for the detection on the Gaussian synthetic DPTV images without image degradation $\mathcal{D}_A$. The resulting effectively maintained degree of precision and recall for this adjustment emphasises that the *Hough* transform requires a fine tuning for each new test dataset, respectively for each new experimental setup. Similarly, the amount of false positive detections for *RetinaNet* can be significantly reduced posteriori, if detections with low certainty scores are excluded from the evaluation after the detection. This can be done at a relatively low cost in the rate of detected particles, while allowing for a significantly higher precision of the detection as can be seen by the sharp drop in the precision-recall diagram in figure 12.

3.4. Unsupervised image-to-image translation for training data refinement

The detection performance of NN-based algorithms trained solely on synthetic PI data and applied to real experimental images is expected to be significantly lower than the performance on synthetic data, since the synthetic training images are not representative for image features on real images [27]. Therefore, the second part of the work focuses on the idea to generate refined synthetic data by employing NNs for image-to-image translation on the synthetic PI dataset $\mathcal{D}_a$ as described in section 2.3. The three unsupervised image-to-image translation methods *MUNIT* [30], *DRIT* [32] and *CycleGAN* [31] were trained on a dataset of two sets of 82.000 unpaired PIs for each the synthetic $\mathcal{D}_a$ and the real PI domain $\mathcal{D}_b$. The NNs of all models were trained for 12 epochs, while early stopping of the training was used to counteract degenerate results from a mode collapse or an overfitting of the NN. The best results were reached after 10 training epochs for *MUNIT*, 6 epochs for *DRIT* and 3 epochs for *CycleGAN*.

The resulting refined synthetic PIs from the image-to-image translation by the best version of each translation algorithm are contrasted in figure 13. Overall the visually best results were reached by *MUNIT*, which was able to reproduce small scale image features leading to realistic synthetic representations of real images. It can be seen that the distribution of synthetic PIs

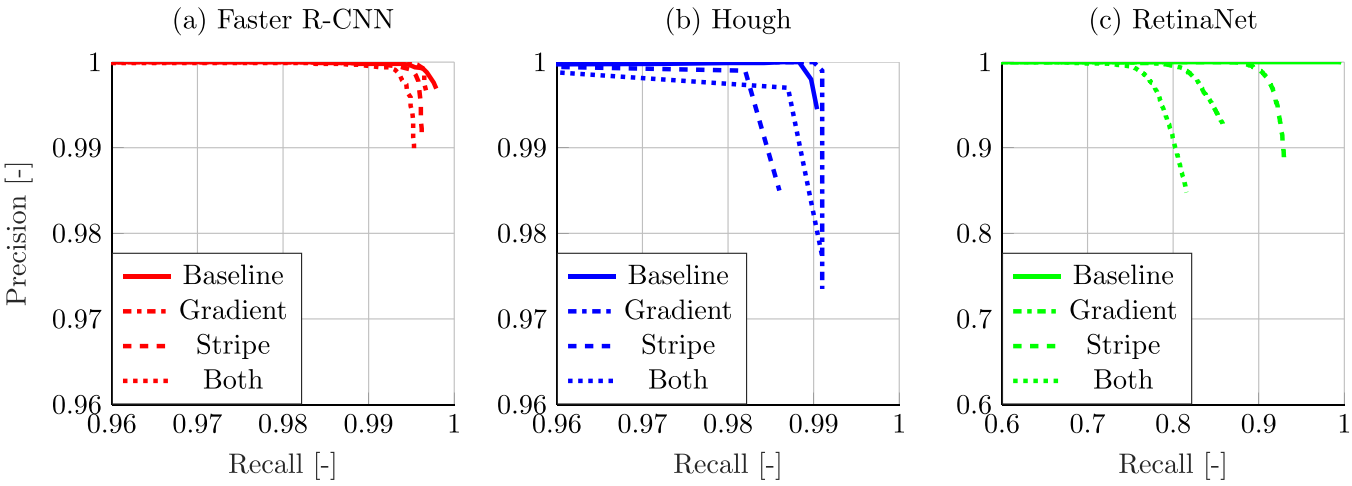
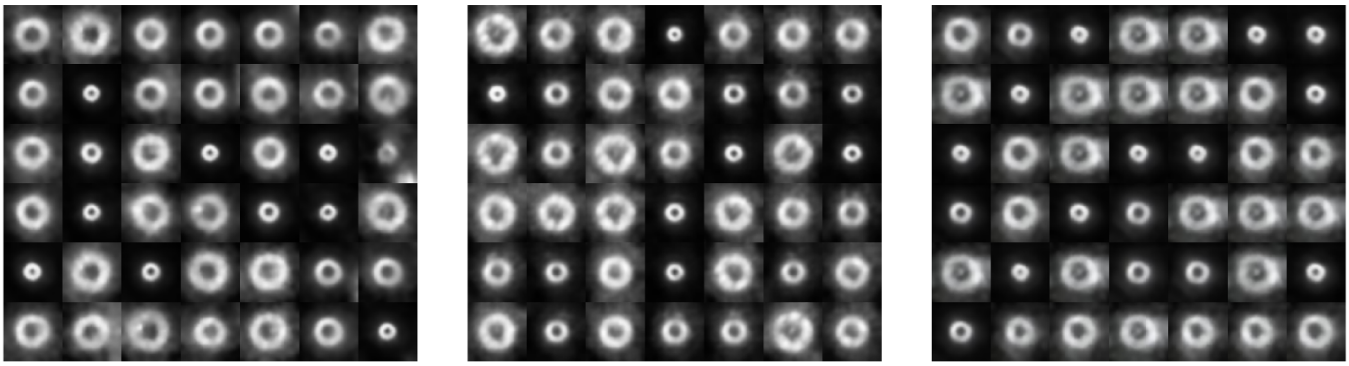


Figure 12. Comparison of the detection performance by the different particle detection approaches on degraded synthetic images with gradients and stripes (axis change for RetinaNet).



(a) *MUNIT* (b) *DRIT* (c) *CycleGAN*

Figure 13. Resulting PIs from training data refinement $\mathcal{D}_c$ with the unsupervised image-to-image translation methods (a) *MUNIT*; (b) *DRIT* and (c) *CycleGAN*.

was transformed closer to the distribution of the real images by the reproduction of increasingly wide edges for higher diameters, noisy contours, distortions and intensity variations over the circumference of the PIs. The PIs generated by *DRIT* did not preserve these features as closely as *MUNIT*, leading to less realistic PIs. As can be seen in figure 13(c) *CycleGAN* suffered from mode-collapse, a well-known failure mode of GANs, in which the generation of a singular mode is exploited by the optimisation scheme in an effort to minimise the objective function [47]. Based on these findings *MUNIT* was selected for the generation of the training data, which is used for the optimisation of the NN-based approaches *Faster R-CNN* and *RetinaNet* for particle detection in the following sections.

The radial intensity profile is a major characteristic of the PIs, as it determines the gradient distribution of the PIs, which is a relevant feature for particle detection. The resulting mean intensity profiles and azimuthal standard deviations for four characteristic diameters from the dataset of refined synthetic PIs $\mathcal{D}_c$ generated with *MUNIT* are also added to figure 3 in order to provide an immediate comparison to the intensity profiles of the Gaussian synthetic images $\mathcal{D}_a$ and real DPTV images $\mathcal{D}_b$. The profile statistics were evaluated over a sample of 100 PIs for each characteristic diameter and the profiles were normalised with the maximum intensity measured over the azimuthal distribution of the radial intensity. As can be seen the radial profile is approximated more closely by *MUNIT* for all characteristic diameters in comparison to the Gaussian synthetic images. Also, the range of intensity variation over the circumference of the PIs is modelled more closely as well by *MUNIT*, which is indicated by the respective standard deviations. The combination of the reproduction of small scale features and a close approximation of the radial intensity profile, therefore, candidates *MUNIT* a particularly promising approach to generate refined synthetic PIs with a high visual similarity to the real PIs observed in the DPTV experiments.

The dataset of PIs used for the training of the image translation NNs contained residual overlapping and distorted PIs in the domain of real PIs, which was gathered by cutting out particles from the experimental images. These outliers were filtered out by the NNs during the image translation step, so that within the refined synthetic dataset $\mathcal{D}_c$ no overlapping or

irregular PIs were generated. This data cleaning effect may result from the exclusion of uncommon modes by the autoencoders, which are part of the NN architecture in the image-to-image translation algorithms [48].

During image translation the diameter of the individual synthetic input images was not preserved in the output PIs, leading to a MAE of $\varepsilon_D = 1.2$ px in PI diameter d_e for the best image-to-image translation model. However, the errors in the planar position ε_X and ε_Y were consistently low for all translation models, with a MAE lower than 0.3 px (see table A3). It was furthermore found that a dataset imbalance in respect to the diameter distribution of the real PIs influences the accuracy of the diameter translation negatively by skewing the translation towards the most common value in the diameter distribution. To overcoming this adverse effect and in turn improve the diameter-preservation accuracy during translation the training dataset was adjusted towards an uniform abundance distribution of PI diameters. Even though the diameter deviation could not be fully corrected, the MAE was reduced to $\varepsilon_D = 0.8$ px with the modified dataset. The remaining MAE, therefore, requires an additional correction step for the recalculation of the diameter label of the generated refined synthetic PIs. The diameter and position was measured by a sub-pixel accurate 5th order polynomial fit, that detected the intensity peak in a 3×11 px kernels at 180 azimuthal positions. The uncertainty of this reference measurement is expected to influence the accuracy of the NN-based detectors trained on this data. It can however be assumed that the NN to some degree will be able to increase their accuracy beyond the training data accuracy, if (i) the label errors are zero mean and (ii) a statistically significant number of labelled PIs is available. In this case the optimisation of the NN over all samples could lead to a compensation of the errors.

While the radial intensity distribution of singular PIs is accurately reproduced, the absolute intensity level of the generated PIs appears to be normalised during image translation, leading to overall brighter PIs, especially for high diameters, as can be seen in figure 13. This effect is considered to be caused by the instance normalisation layers [49], which are part of the CNNs of all image translation frameworks. As such, the intensity level of the refined synthetic PIs has been retroactively retrieved by a correction function that was determined from the linearised evolution of the peak intensity over the diameter observed in real PIs. The corrected translation refined synthetic PIs were then composed to a dataset of refined synthetic DPTV images $\mathcal{D}_C$, which was used to train the NN-based particle detectors with the scope of an enhanced detection rate.

3.5. Particle detection on experimental images

The NN-based particle detection methods *Faster R-CNN* and *RetinaNet* are compared to the *Hough* transform by means of experimental recordings from DPTV experiments [15]. Moreover, the direct comparison between NN-based particle detectors trained on refined synthetic data generated by the proposed method and the same networks trained on real PIs from the original DPTV recordings allows an evaluation of the impact of synthetic training data refinement. Accordingly, the NN-based approaches *Faster R-CNN* and *RetinaNet* are trained as separate versions each on simple Gaussian PIs $\mathcal{D}_a$, refined synthetic images from the *MUNIT* image-to-image-translation $\mathcal{D}_c$ and real cut-out PIs $\mathcal{D}_b$. The training on cut-out PIs instead of complete experimental images annotated by an algorithm was chosen, since the unlabelled particles in the experimental images severely impacted the training of the NNs, as they present a large amount of false negative examples. This effect was confirmed during preliminary testing on experimental images annotated by the *Hough* transform, which leads to a comparably low detection performance. Consequently, a dataset $\mathcal{D}_B$ that resembles the experimental image dataset, but contains only labelled PIs $\mathcal{D}_b$ was generated by cutting-out, measuring and pasting PIs from DPTV recordings on an uniform background.

The comparison of the suitability of the respective PIs as training data is possible, since all datasets $\mathcal{D}_A$, $\mathcal{D}_B$ and $\mathcal{D}_C$ are fully labelled. The NNs of the particle detectors were trained on each $n_t = 49\,000$ images from the respective datasets $\mathcal{D}_A$, $\mathcal{D}_B$ and $\mathcal{D}_C$ and reached an optimal performance after 2 training epochs. Training was repeated five times for both *Faster R-CNN* and *RetinaNet* on each training dataset and the NNs with the highest average precision on the real test dataset were selected. A noticeable variation in the performance of the NN-based particle detectors of 0.02 to 0.07 points of *AP* for *Faster R-CNN* and 0.04 to 0.16 for *RetinaNet* was observed, which is attributed to the random initialisation of the parameters within the NNs.

The test dataset constituted manually annotated DPTV images from the grooved region of the open wet clutch disk used in the experiments from Leister and Kriegseis [15], for which an example can be seen in figure 11. As no ground truth is available for the experimental images, bounding box labels were drawn manually on the intensity maximum at the horizontal and vertical extrema of the PIs, consistent with the training dataset. This way a dataset of $n_e = 50$ experimental recordings with an approximate $n_p = 1000$ labelled PIs in total was created. The manually labelled images can be considered an acceptable ground truth for the detection performance in terms of precision and recall. However, they are not suitable for an analysis of the detection accuracy, due to the introduction of additional bias and uncertainty from manual image annotation. The dynamic range of the 14bit raw images was adjusted to the range of occupied intensity values and the resulting images were saved as 8 bit png files. These images were directly used for evaluation of the NN-based particle detectors, while for the detection with the *Hough* algorithm the images were first further preprocessed with a global minimum subtraction.

The results of the detection by the NN-based detection methods *Faster R-CNN* and *RetinaNet* as well as the model based *Hough* algorithm are summarised in figure 14. Training of the particle detection NNs on the translation-refined synthetic images $\mathcal{D}_C$ (labelled with *MUNIT* in the diagram) leads to a significantly higher detection performance in comparison to simple synthetic images $\mathcal{D}_A$ as well as real cut-out images $\mathcal{D}_B$. The NN-based particle detectors trained on simple synthetic images $\mathcal{D}_A$ yield the lowest relative performance,

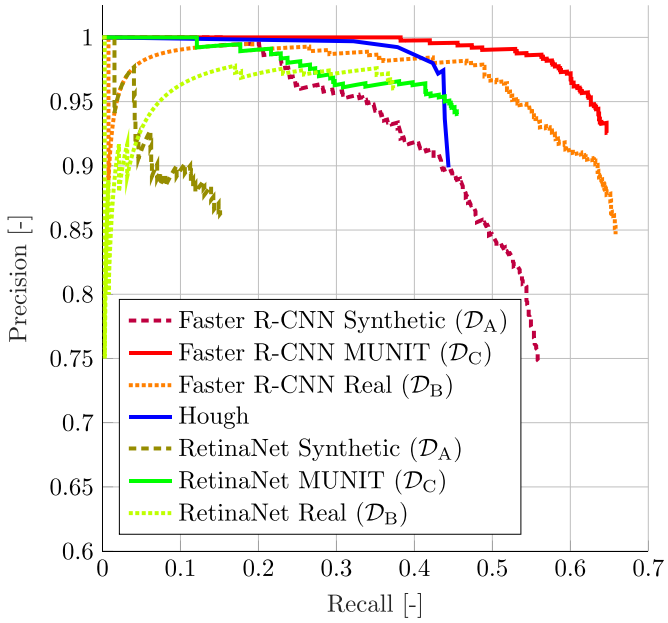


Figure 14. Detection performance on real DPTV images for *Faster R-CNN* and *RetinaNet* trained on the synthetic $\mathcal{D}_A$, translation-refined synthetic $\mathcal{D}_C$ and semi-synthetic $\mathcal{D}_B$ datasets in comparison to the *Hough* algorithm.

which can be explained by a discrepancy in the image features of the synthetic training images to the real test images. The importance of representative training data was already reported by Barnkob *et al* [12] who trained a *Faster R-CNN* detector for particle detection on synthetic APTV data and tested the NNs on synthetic images with additional noise and overlaps. Especially for *RetinaNet* a significant increase in the rate of detected particles can be achieved through synthetic training data refinement, tripling the recall from 0.152 for the NNs trained on simple synthetic data $\mathcal{D}_A$ to 0.456 by training *RetinaNet* on the refined synthetic data $\mathcal{D}_C$. Likewise, the precision is also increased such that *RetinaNet* even surpasses the benchmark *Hough* transform algorithm in both precision and recall by a small margin. This significant improvement highlights the dependence of *RetinaNet* on suitable training data, i.e. representative training images with a high similarity to the test images. *Faster R-CNN* trained on synthetic images $\mathcal{D}_A$ already reaches a higher detection rate than the *Hough* transform on the real test images, however at the cost of a higher amount of false positive detections as indicated by a relatively low precision. Training of *Faster R-CNN* on the refined synthetic data $\mathcal{D}_C$, however, leads to significant gains in precision and the amount of successfully detected particles, leading to a 47.4% higher detection rate in comparison to the *Hough* transform.

Note that NNs were found to be more robust to noise in direct comparison of the NNs and the considered DPTV datasets [15, 17], in the latter has a significantly higher SNR than the former. In the case of a lower SNR the NNs detected a higher amount of particles relative to the *Hough* algorithm.

Furthermore, translation-refined synthetic training data $\mathcal{D}_c$ reveal a higher detection performance of the NN-based particle detection methods in comparison to training images composed of real cut-out PIs $\mathcal{D}_b$. In the case of *Faster R-CNN* this is mainly manifested by a 9.3% improvement in precision, while *RetinaNet* gains a 21.9% improvement in recall through training on the refined dataset. The achieved improvement underlines the high representative value of the translation-refined PIs for real DPTV image features and indicates that this method of training data refinement is strongly beneficial, since it even outperforms real PI training data. This positive effect is assumed to result from the cleaning of the training data, since during the image-to-image translation overlapping and distorted PIs as well as other outliers are not translated onto the refined synthetic PIs as described in section 3.4.

The presented approach to synthetic training data refinement by unsupervised image-to-image translation enables the NN-based particle detection methods to reach a significantly higher detection performance in comparison to the simple synthetic data based on model functions, measured by a combined precision and recall on the real test images in comparison to the simple synthetic training data. This shows that the image feature distribution of the synthetic dataset was successfully moved closer to the real distribution with the proposed refinement approach, leading to more representative training data for the NNs and in turn an improved generalisation and detection on real-world DPTV images. In consequence, this improvement enabled the NN-based approaches to surpass the detection performance of the benchmark *Hough* algorithm for the chosen test dataset of DPTV recordings from the experiments.

4. Discussion

As confirmed by the example of the two considered object detection frameworks *Faster R-CNN* and *RetinaNet*, NNs are a versatile approach for particle detection, offering the potential for improvements in both accuracy and detection rate of DPTV.

The gain in the precision of the NN approach in combination with synthetic training data refinement underlines the importance of suitable training data for NN-based particle detection. The presented method for synthetic training-data refinement allows for a fine tuning of the NNs for a particular experimental setup by learning characteristic features of the defocused PIs, thus significantly enhancing the performance of the NN-based particle detectors for a particular measurement setup. However, in order to improve the generalisation capabilities of the particle detectors on testing data from different experimental setups, in which the features of the PI deviate from the current distribution, a systematic extension of the training distribution is mandatory.

The comparison of *Faster R-CNN* and *RetinaNet* revealed a different dependence on the suitable training data, where *RetinaNet* relies more on representative training data in comparison to *Faster R-CNN*. This insight is evident from the larger improvements in detection rate and precision from training on refined synthetic PI data $\mathcal{D}_c$ (cp. figure 14). Since *RetinaNet* takes advantage of a FPN [37], smaller scale image features are considered in comparison to standard CNN feature-extraction

approaches. While this network architecture has been shown to be beneficial for the detection accuracy, as can also be seen in table 1, the detector happens to be more discriminative, as indicated by the low detection rate with less representative training data. The standard CNN based *Faster R-CNN*, in contrast, demonstrates more robustness to changes in the test images. The comparatively lower measurement accuracy of *Faster R-CNN* further underlines the importance of small-scale image features for an accurate measurement of the particle position.

The training procedure for the presented two-stage NN particle detection method involves a considerable effort for the preparation of the training datasets and additional temporal effort due to the computationally expensive optimisation process of the NNs during training. However, the training of the NNs has to be performed as a pre-processing step only once for a particular measurement setup, which is a clear long-term advantage over the necessary adjustment of most conventional particle detection methods for each new experimental study. The latter e.g. involves iterative adjustments of the *Hough* for an appropriate rejection criterion under varying illumination conditions. For the NN-based detection approaches, in contrast, such adjustment is done *a posteriori* based on certainty scores of the singular particle detections.

The computational time of the NN-based approaches is found to be an order of magnitude higher as compared to the *Hough* algorithm, i.e. 2.7 s (*Faster R-CNN*) and 5.52 s (*RetinaNet*) versus 0.3 s (*Hough*) per image on a common laptop computer. Since the full datasets accordingly are processed in only a few hours and moreover automatically, this effort poses no significant impediment. Also, during the optimisation of the NNs image features that are relevant for the recognition of the PIs characteristics are determined automatically, which reduces the necessary modelling efforts significantly. Therefore, NN-based particle detection can be quickly adapted for new experimental conditions with little prior knowledge of the defocused PI characteristics, thus making DPTV more accessible for inexperienced users.

The advantage of NN-based particle detection over the conventional *Hough* transform becomes more significant when the respectively considered images diverge from the underlying assumptions, as emphasised by the reduced performance on real images compared to synthetic images. Aberrations that are characteristic for a given measurement setup are internalised as features during training of the NN and can, therefore, be utilised beneficially for the detection. As was shown in this work, NNs are also more robust to fluctuations in illumination, which is particular advantageous for measurement setups comprising illumination imperfections and/or reflections. This potential of NN-based particle detection likewise holds true for aberration issues, which accordingly allows for the application standard optical components without loss of information—thus, in turn, increasing the availability of defocusing particle tracking for as wider range of researchers.

5. Conclusions

The present investigation suggests that a two staged approach based on synthetic training data refinement by unsupervised image-to-image translation and object detection leads to significant improvements in particle detection for DPTV.

Particularly, particle detectors trained on the refined synthetic data are shown to reach a significantly higher performance in terms of combined precision and recall in comparison to the same detectors trained on simple Gaussian PIs. This improvement is an important insight, since it immediately emphasises the necessity of representative small-scale image features in the training data for any advanced particle-detection approaches. This requirement has been most saliently shown for*RetinaNet*, which performs detection on higher resolution feature maps and allows for a comparatively high spatial accuracy. Obviously, representative small scale image features are therefore more rigorously required for *RetinaNet* in order to reach a high detection rate, since the employment of higher resolution features is assumed to increase the detectors specificity.

Since the accurate manual annotation of real image data is unfeasible, further development of methods for synthetic training data refinement seems necessary. The above discoveries and insights, consequently, come with the intruding conclusion that further increase in particle detection accuracy can be expected for more accurately labelled training data. That is, an improvement of the training data refinement step towards a better spatial conservation is required, especially for a closer preservation of the PIs diameter. In this context it appears particularly promising to constrain the particle translation of the general purpose image-to-image translation method *MUNIT* with an additional loss term so as to force a more accurate shape preservation. This strategy is envisioned to be realisable by the per-pixel loss of the synthetic input and refined output image. Another approach could be a more recent framework for unsupervised image-to-image translation that uses an additional NN to determine the necessary degree of shape preservation automatically from the training dataset [50].

The NN-based particle detection approach—especially in case of *Faster R-CNN*—is shown to be robust towards illumination variations in the background, a low SNR and blurry PIs resulting e.g. from image aberration. The approach is able to maintain a high detection rate on low quality DPTV images, which proved to be challenging for the conventional *Hough* transform [15, 17]. The proposed approach can be employed to a new experimental setup without prior knowledge of the PI characteristics, since the NNs learn the relevant features of the PIs in an automated procedure from the image data and furthermore utilise image aberration in a beneficial way as features for the detection. This makes the NN-based approach versatile in respect to different experimental conditions, as already demonstrated successfully for different APTV [12, 21, 24], therefore increasing the availability for a wider range of less experienced users.

In contrast to the conventional *Hough* transform-based approach the NN-based particle detection was found to resolve overlapping PIs at substantially higher overlap and with a significantly lower loss in accuracy in comparison to singular PIs. These insights lead to the conclusion that NN-based approaches provide the prospect of a significant increase in particle seeding density in the DPTV experiments. Future investigation might consequently be directed at evaluating the new approach on DPTV experiments with an increased particle seeding density in order to evaluate how well the improvements on synthetic data translate to practical experiments.

A comparison of the detection on synthetic images revealed that the NN-based particle detection approach offers the potential for an increase in measurement accuracy. Especially in diameter determination both *Faster R-CNN* and *RetinaNet* reach a higher accuracy in comparison to the *Hough* transform, which results in a more homogeneous distribution in planar and out-of-plane accuracy. The *RetinaNet* object detector allows for a more significant improvement in localization accuracy, while *Faster R-CNN* offers larger improvements in the detection rate in comparison to the *Hough* algorithm. This insight opens up the possibility to optimise particle detection in DPTV depending on the desired measurement properties. It it therefore concluded that hybrid NN models on the basis of the two evaluated detection methods *Faster R-CNN* and *RetinaNet* seem to be the most promising approach by combining pivotal traits into a custom NN architecture for particle detection. Consequently, an adaption of the FPN to *Faster R-CNN* is expected to increase the accuracy of the particle detection, since the FPN allows for the use of higher resolution features for detection. A second promising hybrid approach is foreseen to result from merging of NNs with conventional particle detection algorithms, in which the NN generates region proposals for a conventional particle measurement algorithm, thus combining the high detection rate of the NN based detection with the high in-plane accuracy of conventional methods. Such a hybrid approach might be directly realised on the basis of *Faster R-CNN* and the particle position refinement step of the *Hough* algorithm.

Even though NN-based particle detection approaches are rather recent developments in comparison to established conventional detection methods based on model functions and cross-correlation, a competitive performance to the *Hough* transform is achieved with NN-based approaches for all presented experimental conditions. This in turn indicates further improvements to be expected in particle tracking by means of NNs. Such expectations—in context of the present study—lead to the following final remarks.

A variety of general-purpose object-detection NN architectures are available and can be straight forwardly adopted for particle detection. Further specialised CNNs, especially for the detection of small objects are envisioned to offer additional potential performance gains beyond the current approach. Promising candidates are perceptual GANs for small object detection [51], based on the super-resolution of small objects, feature-fused single shot detectors [52] that utilise additional contextual information and object detectors with scale-dependant pooling [53].

The development of physics-informed NNs dedicated for particle tracking seems likewise worthwhile. The temporal information contained in the pathlines of the particles, for instance, might be exploited by a LSTM network to guide the training procedure and might, therefore, allow for a more accurate measurement, by excluding nonphysical detections.

Finally, training procedures in particular for PTV have to be further developed in order to better exploit the capability of the NNs, which shifts the focus towards the provision of suitable training data for the optimisation of the NNs. Due to the need of large datasets for training the networks, synthetic data refinement as proposed and successfully tested in the present work is expected to play an important role in improving the NN-based perspective on particle detection approaches.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5445/IR/1000146837.

Appendix

Table A1. Training settings of the particle detection methods, *Faster R-CNN* with InceptionV2 base network and *RetinaNet* with ResNet50 base network.

Parameter	*Faster R-CNN*	*RetinaNet*
Optimizer	SGDM	SGDM
Epochs	2–5	2–5
Minibatch size	4	4
Learning rate α	0.0002	0.04
Learning rate decay	None	cosine
Warm up period	None	2000
Momentum β	0.9	0.9
L_2 regularization	None	0.0004

Table A2. Training settings of the image-to-image translation methods *CycleGAN*, *MUNIT* and *DRIT*.

Parameter	*CycleGAN*	*MUNIT*	*DRIT*
Optimizer	Adam	Adam	Adam
Epochs	3	10	6
Minibatch	1	1	1
lr α	0.0002	0.0001	0.0001
lr decay	Stepwise[a]	Stepwise[b]	Linear
β_1	0.5	0.5	0.5
β_2	0.999	0.999	0.999
L_2 reg	none	0.0001	0.0001

[a] Learning rate halved every 200 000 training iterations.
[b] Learning rate halved every 100 000 training iterations.

Table A3. MAEs in planar position and diameter during the image-to-image translation with *CycleGAN*, *MUNIT* and *DRIT*.

Parameter	*CycleGAN*	*MUNIT*	*DRIT*
ε_D (px)	1.187	1.823	**0.832**
ε_X (px)	0.187	**0.159**	0.286
ε_Y (px)	**0.146**	0.181	0.307

ORCID iDs

Maximilian Dreisbach https://orcid.org/0000-0001-6308-0982
Robin Leister https://orcid.org/0000-0002-0286-8183
Matthias Probst https://orcid.org/0000-0001-8729-0482
Pascal Friederich https://orcid.org/0000-0003-4465-1465
Alexander Stroh https://orcid.org/0000-0003-0850-9883
Jochen Kriegseis https://orcid.org/0000-0002-2737-2539

References

[1] Raffel M, Willert C, Scarano F, Kähler C J, Wereley S T and Kompenhans J 2018 *Particle Image Velocimetry—A Practical Guide* (Berlin: Springer)
[2] Chang T, Watson A T and Tatterson G B 1985 Image processing of tracer particle motions as applied to mixing and turbulent flow—I. The technique *Chem. Eng. Sci.* **40** 269–75
[3] Wu M, Roberts J, Kim S, Koch D and DeLisa M 2006 Collective bacterial dynamics revealed using a three-dimensional population-scale defocused particle tracking technique *Appl. Environ. Microbiol.* **72** 4987–94
[4] Nishino K, Kasagi N and Hirata M 1989 Three-dimensional particle tracking velocimetry based on automated digital image processing *J. Fluids Eng.* **111** 384–91
[5] Maas H G, Gruen A and Papantoniou D 1993 Particle tracking velocimetry in three-dimensional flows *Exp. Fluids* **15** 133–46
[6] Schanz D, Schröder A, Gesemann S, Michaelis D and Wieneke B 2013 'Shake The Box': a highly efficient and accurate tomographic particle tracking velocimetry (TOMO-PTV) method using prediction of particle positions *10th Int. Symp. on Particle Image Velocimetry (Delft, The Netherlands, 1–3 July 2013)* pp 1–13 (available at: http://resolver.tudelft.nl/uuid:212b0c2d-3210-482f-b751-91d98d5ea43d)
[7] Kao H and Verkman A 1994 Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position *Biophys. J.* **67** 1291–300
[8] Willert C E and Gharib M 1992 Three-dimensional particle imaging with a single camera *Exp. Fluids* **12** 353–8
[9] Pereira F J A, Lu J, Castaño-Graff E and Gharib M 2007 Microscale 3D flow mapping with μ DDPIV *Exp. Fluids* **42** 589–99
[10] Wu M, Roberts J and Buckley M 2005 Three-dimensional fluorescent particle tracking at micron-scale using a single camera *Exp. Fluids* **38** 461–5
[11] Olsen M and Adrian R 2000 Out-of-focus effects on particle image visibility and correlation in microscopic particle image velocimetry *Exp. Fluids* **29** S166–74
[12] Barnkob R, Cierpka C, Chen M, Sachs S, Mäder P and Rossi M 2021 Defocus particle tracking: a comparison of methods based on model functions, cross-correlation and neural networks *Meas. Sci. Technol.* **32** 094011
[13] Adrian R J and Yao C-S 1985 Pulsed laser technique application to liquid and gaseous flows and the scattering power of seed materials *Appl. Opt.* **24** 44–52
[14] Fuchs T, Hain R and Kähler C 2016 *In situ* calibrated defocusing PTV for wall-bounded measurement volumes *Meas. Sci. Technol.* **27** 084005
[15] Leister R and Kriegseis J 2019 3D-LIF experiments in an open wet clutch by means of defocusing PTV *13th Int. Symp. on Particle Image Velocimetry (ISPIV 2019) (Munich, Germany, 22–24 July 2019)* ed C J Kähler, R Hain, S Scharnowski and T Fuchs (https://doi.org/10.5445/IR/1000098119)
[16] Rhody H 2005 Lecture 10: Hough circle transform (Chester F. Carlson Center For Imaging Science, Rochester Institute of Technology) (available at: https://www.cis.rit.edu/class/simg782/lectures/lecture_10/lec782_05_10.pdf)
[17] Leister R, Fuchs T, Mattern P and Kriegseis J 2021 Flow-structure identification in a radially grooved open wet clutch by means of defocusing particle tracking velocimetry *Exp. Fluids* **62** 29
[18] Cierpka C, Segura R, Hain R and Kähler C J 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** 045401
[19] Barnkob R, Kähler C J and Rossi M 2015 General defocusing particle tracking *Lab Chip* **15** 3556–60
[20] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
[21] Cierpka C, König J, Chen M, Boho D and Mäder P 2019 On the use of machine learning algorithms for the calibration of

astigmatism PTV *13th Int. Symp. on Particle Image Velocimetry (ISPIV 2019) (Munich, Germany, 22–24 July 2019)* ed C J Kähler, R Hain, S Scharnowski and T Fuchs (available at: https://athene-forschung.unibw.de/129121?id=129121)
[22] Ren S, He K, Girshick R and Sun J 2017 Faster R-CNN: towards real-time object detection with region proposal networks *IEEE Trans. Pattern Anal. Mach. Intell.* **39** 1137–49
[23] LeCun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** 436–44
[24] König J, Chen M, Rösing W, Boho D, Mäder P and Cierpka C 2020 On the use of a cascaded convolutional neural network for three-dimensional flow measurements using astigmatic PTV *Meas. Sci. Technol.* **31** 074015
[25] Franchini S and Krevor S 2020 Cut, overlap and locate: a deep learning approach for the 3D localization of particles in astigmatic optical setups *Exp. Fluids* **61** 140
[26] Stewart R, Andriluka M and Ng A Y 2016 End-to-end people detection in crowded scenes *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV, USA, 27–30 June 2016)* pp 2325–33
[27] Shrivastava A, Pfister T, Tuzel O, Susskind J, Wang W and Webb R 2017 Learning from simulated and unsupervised images through adversarial training *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 2242–51
[28] Dreisbach M, Leister R, Probst M, Friederich P, Stroh A and Kriegseis J 2022 Particle Detection by means of Neural Networks and Synthetic Training Data Refinement in Defocusing Particle Tracking Velocimetry (data) (https://doi.org/10.5445/IR/1000146837)
[29] Rossi M 2020 Synthetic image generator for defocusing and astigmatic PIV/PTV *Meas. Sci. Technol.* **31** 017003
[30] Huang X, Liu M Y, Belongie S and Kautz J 2018 Multimodal unsupervised image-to-image translation *Computer Vision—ECCV 2018* (Cham: Springer International Publishing) pp 179–96
[31] Zhu J, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *2017 IEEE Int. Conf. on Computer Vision (ICCV) (Venice, Italy, 22–29 October)* pp 2242–51
[32] Lee H Y, Tseng H Y, Huang J B, Singh M and Yang M H 2018 Diverse image-to-image translation via disentangled representations *Computer Vision – Eccv 2018*, ed V Ferrari, M Hebert, C Sminchisescu and Y Weiss (Cham: Springer International Publishing) pp 36–52
[33] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative adversarial nets *Advances in Neural Information Processing Systems* vol 27 (Curran Associates, Inc.) pp 2672–80 (available at: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf)
[34] Rumelhart D E, Hinton G E and Williams R J 1986 *Learning Internal Representations by Error Propagation* (Cambridge, MA: MIT Press) pp 318–62
[35] Lin T-Y, Goyal P, Girshick R, He K and Dollár P 2020 Focal loss for dense object detection *IEEE Trans. Pattern Anal. Mach. Intell.* **42** 318–27
[36] Redmon J and Farhadi A 2018 YOLOv3: an incremental improvement (arXiv:1804.02767)
[37] Lin T Y, Dollár P, Girshick R, He K, Hariharan B and Belongie S 2017 Feature pyramid networks for object detection *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (Los Alamitos, CA: IEEE Computer Society) pp 936–44
[38] Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z and Qu R 2019 A survey of deep learning-based object detection *IEEE Access* **7** 128837–68
[39] Abadi M *et al* 2015 TensorFlow: large-scale machine learning on heterogeneous systems (arXiv:1603.04467)
[40] Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L 2014 Microsoft COCO: common objects in context *Computer Vision–Eccv 2014* ed D Fleet, T Pajdla, B Schiele and T Tuytelaars (Cham: Springer) pp 740–55
[41] Qian N 1999 On the momentum term in gradient descent learning algorithms *Neural Netw.* **12** 145–51
[42] Kingma D and Ba J 2015 Adam: a method for stochastic optimization *3rd Int. Conf. on Learning Representations, Conf. Track Proc. (ICLR 2015) (San Diego, CA, USA)*
[43] Bendat J S and Piersol A G 2010 *Random Data: Analysis and Measurement Procedures (Wiley Series in Probability and Statistics)* 4th edn (New York: Wiley) (available at: www.wiley.com/en-gb/Random+Data:+Analysis+and+Measurement+Procedures,+4th+edn-p-9780470248775
[44] Sammut C and Webb G I 2011 *Encyclopedia of Machine Learning* 1st edn (New York: Springer) (https://doi.org/10.1007/978-0-387-30164-8)
[45] Manning C D, Raghavan P and Schütze H 2008 *Introduction to Information Retrieval* (Cambridge: Cambridge University Press)
[46] Everingham M, Van Gool L, Williams C K I, Winn J and Zisserman A 2010 The pascal visual object classes (VOC) challenge *Int. J. Comput. Vis.* **88** 303–38
[47] Goodfellow I 2016 Nips 2016 tutorial: generative adversarial networks (arXiv:1701.00160)
[48] Vincent P, Larochelle H, Bengio Y and Manzagol P A 2008 Extracting and composing robust features with denoising autoencoders *Proc. 25th Int. Conf. on Machine Learning (ICML '08)* (New York: Association for Computing Machiner) pp 1096–103
[49] Ulyanov D, Vedaldi A and Lempitsky V S 2017 Improved texture networks: maximizing quality and diversity in feed-forward stylization and texture synthesis *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 4105–13
[50] Kim J, Kim M, Kang H and Lee K 2020 U-GAT-IT: unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation (arXiv:1907.10830)
[51] Li J, Liang X, Wei Y, Xu T, Feng J and Yan S 2017 Perceptual generative adversarial networks for small object detection *2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 1951–9
[52] Cao G, Xie X, Yang W, Liao Q, Shi G and Wu J 2018 Feature-fused SSD: fast detection for small objects *Proc. SPIE* **10615** 381–8
[53] Yang F, Choi W and Lin Y 2016 Exploit all the layers: fast and accurate CNN object detector with scale dependent pooling and cascaded rejection classifiers *2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 2129–37