

PAPER • OPEN ACCESS

Deep learning and hybrid approach for particle detection in defocusing particle tracking velocimetry

To cite this article: Christian Sax *et al* 2023 *Meas. Sci. Technol.* **34** 095909

View the [article online](#) for updates and enhancements.

You may also like

- [A spatial-temporal algorithm for three-dimensional particle tracking velocimetry using two-view systems](#)
Y Wu, Q Wang and C Y Zhao
- [A review of progress in the physics of open quantum systems: theory and experiment](#)
I Rotter and J P Bird
- [Enhanced Performance of Poly\(3-hexylthiophenes\) Based Thin Film Transistors Using Double-Coated Active Layer](#)
Chia-Hao Chang and Chao-Hsin Chien

Deep learning and hybrid approach for particle detection in defocusing particle tracking velocimetry

Christian Sax* , Maximilian Dreisbach , Robin Leister 
and Jochen Kriegseis 

Institute of Fluids Mechanics (ISTM), Karlsruhe Institute of Technology (KIT), Kaiserstraße 10, 76131 Karlsruhe, Germany

E-mail: christian.sax@kit.edu

Received 30 December 2022, revised 26 March 2023

Accepted for publication 13 April 2023

Published 29 June 2023



CrossMark

Abstract

The present work aims at the improvement of particle detection in defocusing particle tracking velocimetry (DPTV) by means of a novel hybrid approach. Two deep learning approaches, namely faster R-CNN and RetinaNet are compared to the performance of two benchmark conventional image processing algorithms for DPTV. For the development of a hybrid approach with improved performance, the different detection approaches are evaluated on synthetic and images from an actual DPTV experiment. First, the performance under the influence of noise, overlaps, seeding density and optical aberrations is discussed and consequently advantages of neural networks over conventional image processing algorithms for image processing in DPTV are derived. Furthermore, current limitations of the application of neural networks for DPTV are pointed out and their origin is elaborated. It shows that neural networks have a better detection capability but suffer from low positional accuracy when locating particles. Finally, a novel *Hybrid Approach* is proposed, which uses a neural network for particle detection and passes the prediction onto a conventional refinement algorithm for better position accuracy. A third step is implemented to additionally eliminate false predictions by the network based on a subsequent rejection criterion. The novel approach improves the powerful detection performance of neural networks while maintaining the high position accuracy of conventional algorithms, combining the advantages of both approaches.

Keywords: defocusing particle tracking velocimetry, image processing, particle detection, deep learning, neural network, CNN, hybrid-approach

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

* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

1. Introduction

Particle imaging techniques are a popular group of non-intrusive optical measuring techniques within fluid dynamics. With increasing capability of digital image processing, methods like particle imaging velocimetry (PIV) [1] can provide quantitative field velocity information while having reasonable processing times and uncertainty margins. While in PIV the displacement of an ensemble of particles is correlated, in particle tracking velocimetry (PTV) the particles are tracked individually, enabling a Lagrangian frame of reference and, therefore, providing more physical insight into the flow field. However, planar PIV and PTV are not suited for measurements of a three dimensional flow topology, since only particle displacements within the light sheet are visible. Consequently volumetric methods have been developed to gather three dimensional three components (3D-3C) velocity data. Examples are holographic PTV [2], 3D-PTV [3], tomographic PIV [4] and the shake-the-box approaches [5, 6], all of which require multiple cameras. An alternative to multiple camera approaches are single camera methods like defocusing PTV (DPTV) [7, 8] and astigmatism PTV (APT) [9], yielding the advantage of requiring only a single optical access. Both DPTV and APTV obtain three-dimensional information of a particle position by deliberately defocusing the particle image. In DPTV a defocused particle image appears as a defocused ring on the image plane, whose diameter is directly linked to the corresponding particle distance from the focal plane [10] according to

$$d_i^2 = \underbrace{M^2 d_p^2}_{\text{geom. image}} + \underbrace{5.95(M+1)^2 \lambda^2 f_{\#}^2}_{\text{diffraction}} + \underbrace{\frac{M^2 z^{*2} D_a^2}{(s_0 + z^*)^2}}_{\text{defocusing}}, \quad (1)$$

see also figure 1.

Equation (1) can be divided into three terms: the first term describes the geometric image and is composed of the magnification M and the particle diameter d_p . The second term represents the effect of diffraction, containing the wave length of the light λ scattered by the particles and the focal length of the lens $f_{\#}$. The last term characterizes the influence of defocusing the particle image, depending on the distance to the focal plane z^* , the aperture diameter D_a and the distance of the imaging optics to the focal plane s_0 . For a fixed measurement setup all terms describing the particle image diameter except z^* become constant. For microscopic applications the assumption $s_0 \gg z^*$ applies and equation (1) can be accordingly simplified to

$$d_i \propto \left(\text{const.} + z^{*2} \right)^{\frac{1}{2}}, \quad (2)$$

as the defocusing term predominates the problem. With sufficient distance from the focal plane the hyperbolic relation of equation (2) can be further approximated as linear, i.e. $d_i \propto z^*$ [8]. For simplicity the z^* position is normalized to $z = z^*/h$ with h being the depth of the measurement volume.

In DPTV the in-plane position (x, y) of a particle can be simply determined from the center coordinates of the corresponding imaged defocus ring in the image plane (particle image), while the out-of-plane (depth or z) position is obtained from the ring diameter as previously described. Consequently, algorithms suited for DPTV image processing need to be able to reliably determine the particle images' center position as well as their diameter. One straight-forward approach used by Leister *et al* [11, 12] is the coherent Hough transform (CHT) [13–15]—a variation of the Hough transform [16]—which is a gradient based voting algorithm. Fuchs *et al* [8] used an algorithm detecting the edges of the defocus rings by an adaptive threshold applied on the intensity distribution. While the first two methods determine the defocus rings diameter and center position directly from the particle image, Barnkob *et al* [17–19] used cross correlation to compare the measured particle images to reference ones from a calibration stack, both for particle detection and z -position refinement.

With the rise of machine learning in computer vision and neural networks in particular, new methods for particle detection based on convolutional neural networks (CNN) [20, 21] emerged in the DPTV/APTV community. Cierpka *et al* [22] demonstrated the applicability of *Faster R-CNN* [23] for particle detection in APTV. Franchini and Krevor [24] demonstrated an improved detection rate on overlapping particle images for a CNN-based model in APTV. König *et al* [25] used a cascaded CNN on the basis of faster R-CNN on APTV, which showed to have lower position uncertainties on particles with astigmatism and noise compared to conventional algorithms. Barnkob *et al* [26] also used a cascaded version of a CNN for particle detection in DPTV and APTV, which was composed of a preliminary network (faster R-CNN) for locating particles within the image and a second CNN to refine the out-of-plane position. More recently, Dreisbach *et al* [27] applied CNN-based multi-stage (faster R-CNN) and single-stage detectors (RetinaNet [28]) for DPTV, and furthermore analyzed the effect of synthetic training data refinement by means of generative adversarial networks on the network performance for DPTV image processing. It should be noted that—since APTV and DPTV are closely related measuring techniques—findings on the image processing can be expected to be transferable.

Neural networks yield the potential to improve image processing in DPTV as they have shown to outperform conventional algorithms for highly overlapping and astigmatic particle images [24, 25, 27]. However, the deep learning approaches still face some drawbacks to conventional approaches, which are mainly comprised of limited spatial accuracy [26]. In this work, a novel hybrid approach is presented to improve particle detection in DPTV, in order to take combined advantage of the respective strengths of either approaches. The general idea of combining different algorithms in a multi-step algorithm design is not a new method per se, see for example multistage detectors like faster R-CNN or the combination of the Hough transform as a primer for a neural network classifier by D'Orazio *et al* [29].

However, the novelty of the present approach is to combine a neural-network object-detection framework with a conventional algorithm in a targeted order, tailored for particle detection, to optimally exploit the benefits of both methods. This way the trade-off between neural networks and conventional algorithms is effectively eliminated, since the combination merges the advantages of either processing strategy, while excluding the respective weaknesses and/or shortcomings.

To construct a meaningful hybrid approach first the trade-off between neural networks and conventional algorithms need to be better understood. Therefore—and to build upon previous efforts—the performance on typical limitations of conventional detection algorithms is compared between the different detection approaches. Such limitations are e.g. high noise levels, optical aberrations, strongly overlapping particle images and consequently high seeding densities. This way different possibilities to improve image processing in DPTV by using deep learning can be allocated. In analogy to Dreisbach *et al* [27] the two state-of-the-art neural networks RetinaNet [28] (single stage detector) and faster R-CNN [23] (multi-stage detector) are chosen to capture both single-stage and multi-stage detectors. For conventional detection algorithms representing the current benchmark, the distinction proposed by Barnkob *et al* [26] is used. Therefore, the CHT (model function) and the *DefocusTracker*-Software [30] (cross-correlation method) were selected. Finally, after the comparison, the novel hybrid approach is derived and compared to both neural networks and the conventional algorithms.

2. Methodology

2.1. Image acquisition

In order to evaluate the detector performance on real DPTV images, overall 30 test images were chosen from an experiment on an open wet clutch as described by Leister *et al* [12]. The measurement setup consisted of a Quantel Evergreen Nd:YAG laser ($\lambda = 532$ nm, 200 mJ/pulse) illuminating fluorescent particles ($d_{\text{mean}} = 9.84$ μm , $\lambda_{\text{em}} = 584$ nm). The images were recorded in a double frame mode with a PCO.edge 5.5 sCMOS camera (2560×2160 px, 16 bit) equipped with a Questar QM 100 lens. The images were cut to a 600×600 pixel format without other modifications to avoid downsizing artifacts when being fed into the neural networks, resulting in a particle count of 20–30 particles per image with the particle-images diameter ranging from 15 to 31 pixels. Afterwards, the images were pre-processed by mean-image subtraction and additionally an intensity amplification by a fixed factor was applied to improve visibility of the particle images. Finally, the pre-processed images were labeled by a human annotator (by means of maximum radial intensity position to define the particle image edge) to obtain ground-truth positions and diameters for all particle images.

Since ground-truth acquisition for real images is time consuming and limited to the (manual) labeling accuracy,

additionally synthetic images were generated, enabling a more efficient method to obtain large test- and training-data sets. Furthermore, with synthetic images the particle-image positions can be set, to deliberately create overlaps and the signal-to-noise ratio (SNR) can be fully modified. For the image generation *MicroSIG* [30] was used. The settings of *MicroSIG* were chosen in such a way that the set of experimentally acquired images are mimicked in terms of particle-image size, radial intensity distribution, SNR and general optical appearance. More details on the chosen settings are provided in the appendix (table 2). A comparison of a synthetic and experimental particle image is shown in figure 1.

2.2. Training of the neural networks

Training neural networks, especially deep ones, is a non-trivial task as the training procedure is a high-dimensional optimization problem [31] and can influence the networks performance significantly. To allow for a good comparability, the chosen versions of faster R-CNN and RetinaNet (SSD-ResNet50 FPN) build upon the same backbone network for feature extraction (from the images), i.e. ResNet50-backbone [32]. Both networks were trained in the Tensorflow [33] framework. High capacity networks like faster R-CNN and RetinaNet are typically trained on sufficiently large data sets with a lot of variation in classes, their features (such as e.g. texture, posture, size) and the background. DPTV images on the other hand contain relatively little information compared to complex data sets like the frequently used Microsoft-COCO data set [34], for instance, as only the two classes *Particle* and *Background* do exist.

Furthermore, the DPTV images contain only a very limited number of features and little variation compared to conventional computer vision tasks. Therefore, when training high capacity networks on only comparably low information DPTV data, the risk of over-fitting has to be considered carefully, as already discussed e.g. by Barnkob *et al* [26]. Over-fitting occurs when a network starts to memorize the training data and, therefore, loses its capability to generalize to new unknown data [31]. As a consequence, networks are used, which were already pre-trained on a large non-domain-specific data set (namely Microsoft COCO). During the training process the networks learn low-level features like edges, corners etc but also high-level features like the intensity distribution of a particle image or that the ring shape is related to a particle image. These pre-trained networks, therefore, had already learned low-level features and generic higher complexity features like textures and contours. The networks were then trained on the smaller domain-specific data set to learn DPTV specific features. Domain-specific transfer learning from a large general data set has proven to be very efficient and also yields the advantage that smaller domain-specific data sets are sufficient to learn the desired features, since basic features (such as e.g. edges, gradients, shapes, texture) have already been learned. Consequently, the optimization of the neural network is substantially accelerated compared to

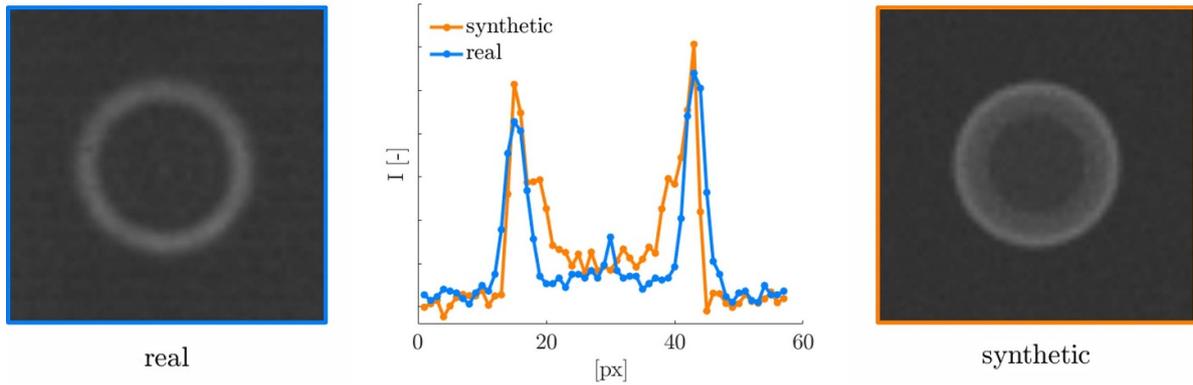


Figure 1. Experimental (real) pre-processed particle image (left) and synthetic particle image (right); the normalized radial intensity distribution of both particle images is shown in the middle.

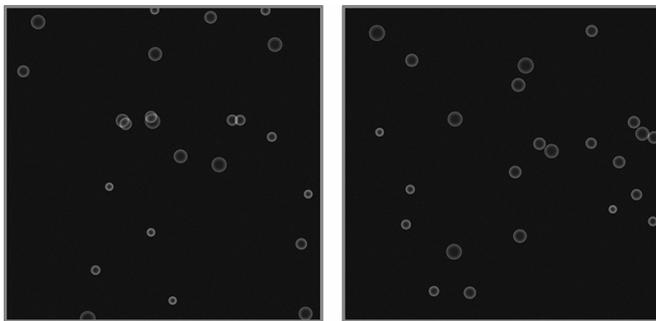


Figure 2. Example images from the synthetic data set used to evaluate the detection algorithms, which were modeled to mimic DPTV images from a real experiment. Images were created with randomly generated particle positions to mimic overlaps like in real DPTV images.

random initialization for training from zero, as only a few neurons (specifically the new output neurons and later layers) have to be adjusted significantly [35–37].

The smaller domain-specific data set consists of 20 000 synthetic DPTV images generated with the *MicroSIG* [30] software as described in section 2.1, see also figure 2. Similar to the real images, the synthetic DPTV images used for training had a range of 20–30 particle images per image, which were randomly distributed within the image. The images had an image resolution of 600×600 pixels to avoid compression effects when being fed into the network. For Faster R-CNN 50 000 and for RetinaNet 25 000 training iterations with stochastic gradient descent with momentum were sufficient for the training losses to converge. A validation data set of 2000 extra images generated in the same manner as the training images were used to detect possible over-fitting. The generation of the training images has been described in greater detail in section 2.1. More details on the training are furthermore given in the appendix (table 1). While the Hough transform does not need any prior knowledge on particle images, it has to be noted that for a significant comparison also synthetic particle images were used to create the calibration stack of the *DefocusTracker*-Software, such that all algorithms build upon the same type of images.

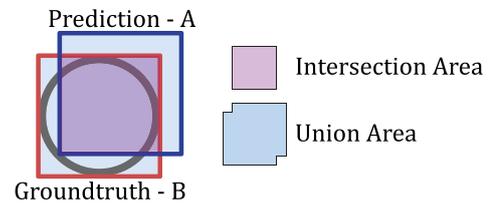


Figure 3. Ground-truth and prediction bounding boxes visualized on a ring shape; the relevant areas for TP and FP determinations—i.e. intersection and union—are furthermore emphasized for clarity.

2.3. Evaluation of the particle detection

For the empirical investigation in this work, the evaluation of the particle-detection process is divided into two categories: first, the general identification of a particle image (for simplification further referred to as just *particle detection* and elaborated below) and, second, the accuracy of the obtained three-dimensional particle-image location estimation (further referred to as *position accuracy* and addressed in section 2.4).

First, a definition of a true positive (TP, correctly detected particle), a false positive (FP, no ground-truth particle—false detection) and a false negative (FN, missed particle) detection must be found. In the field of machine learning and image processing the intersection over union (IoU) is popularly used for this distinction, as depicted in figure 3. A detection is considered to be a TP if the overlap defined by the IoU of the area A of the prediction bounding box with the area B of the ground-truth bounding box, i.e.

$$\text{IoU}_{(A,B)} = \frac{\text{area}(A \cap B)}{\text{area}(A \cup B)}, \quad (3)$$

is above a certain threshold. In this work the commonly used IoU-threshold of 0.5 [23] is used to distinguish a TP ($\text{IoU} \geq 0.5$) from a FP ($\text{IoU} \leq 0.5$), unless stated otherwise. The metrics *precision* and *recall* (PR) are used to rate the particle-detection performance of each detection algorithm, based on the widely spread practice in the field of information retrieval and object detection [38]. Precision and recall are calculated from the number of TP, FP and FN according to

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (4)$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (5)$$

respectively.

The precision ranges from 0 to 1 and rates the capability of an algorithm to have a correct prediction, whereas the recall (also ranging from 0 to 1) rates the amount of particles that were detected by the detector. The evaluation by means of PR allows for a comparison of the algorithms independent from the chosen rejection criterion of each algorithm, since the rejection criterion is systematically varied to achieve a recall from 0 to 1. This is critical, since the detection performance of an algorithm can be heavily impaired by a poor selection of the rejection threshold. Particularly, the rejection for each method is the sensitivity of the Hough transform, the cross-correlation threshold for the cross-correlation method and the uncertainty score outputted by the neural networks, respectively. Note that the rejection criterion restricts the detections of the Hough transform in the initial detection step. For the neural networks, in contrast, a prescribed and accordingly fixed number of outputs is revealed, where the meaningful detections are identified only afterwards through evaluation of the respective confidence scores. That is, the rejection threshold for this extraction process is selected retroactively during the above-mentioned PR evaluation.

Depending on the applied case the required rejection criterion for the optimal PR trade-off can vary. For a rejection criterion chosen too strict, the algorithm achieves high precision but unnecessarily rejects particles, thus leading to low recall values. For the opposite case, a rejection criterion chosen to liberally leads to high recall values as most detections are accepted, but at the cost of low precision, since more FPs are detected. Therefore, the evaluation by means of PR curves bypasses the need for optimal fine tuning of the rejection criterion for each given case, as a wide range of rejection criteria is considered. Additionally the response of an algorithms to a changing rejection criterion can be analyzed.

Note however that both PR depend on the chosen criterion to distinguish TP, FP and FN. Therefore, a quantitative comparison of PR is only possible for identical IoU-thresholds. Precision-recall curves can then be used to compare the quality of the particle detection between the different methods. A good detector is able to achieve a high precision at high recall values, where either precision or recall are prioritized during optimization according to the respective target application.

Another important metric is the average precision (AP), which integrates the precision over the recall via

$$\text{AP} = \int_0^1 P_{(R)} dR \quad (6)$$

and reveals the area under the PR curve. The AP can, therefore, be used to summarize the PR-curve into a single numerical value ranging from zero to one (with one being a perfect detector), which renders AP a particularly convenient metric

to evaluate the performance—especially when an additional third dimension like SNR, overlap or seeding density is added to the analysis.

2.4. Evaluation of the position accuracy

Since the accuracy of DPTV in the image plane (x, y) generally differs from the position accuracy in the depth/out-of-plane position (z , diameter determination) [27], the in-plane and out-of-plane errors will be evaluated separately. As with the precision, the errors are also analyzed over the recall in order to be independent from the selected rejection threshold.

The position errors in x and y direction have been tested to be uncorrelated and underlie a Gaussian distribution. Therefore, the x - and y -position errors can be summed up in the in-plane error

$$\text{Err}_{\text{IP}} = \sqrt{\text{Err}_x^2 + \text{Err}_y^2}, \quad (7)$$

which represents the in-plane length (i.e. magnitude) of the error vector. The probability density function (PDF) of the vector length Err_{IP} , however, is heavily skewed and is found to resemble a Rayleigh distribution

$$f_{(\text{Err})} = \frac{\text{Err}}{\sigma^2} \exp\left[-\frac{\text{Err}^2}{2\sigma^2}\right] \quad (8)$$

rather than a Gaussian one, where σ appears as scale parameter in equation (8), which determines the peak value of the underlying error PDF. Consequently, to provide statistically meaningful metrics for the evaluation of the PDF, this scale parameter σ is determined to describe the characteristic error and additionally the 90th percentile of the corresponding cumulative density distribution (CDF) of the error distribution is chosen to outline the statistical variation of the individual errors. The in-plane errors are indicated in pixels, as an indication in physical coordinates would change depending on the physical measurement setup and would therefore not exclusively rely on the used detection algorithm.

The various detectors use different reference points for the diameter determination of the particle image (e.g. inner vs. outer rim of the ring), since there is no distinct edge but rather a continuous radial intensity distribution (see figure 1). Therefore, a quantitative comparison by diameter is not useful, resulting in a more complex evaluation. Hence, the predicted z positions were plotted over the ground-truth z positions, which should ideally result in a single straight line, with the zero crossing to change depending on the chosen diameter criterion.

For a real detector the predictions will scatter around this line due to measurement errors. To measure the z -position error the scattering points were fitted with a linear fit function using the least square method. Subsequently, the PDF of the absolute of the individual residuals was evaluated according to $\text{Err}_{z(i)} = |\text{residual}_{(i)}^{\text{lin fit}}|$. This PDF describes an absolute value, which similarly follows a Rayleigh distribution. Consequently, as elaborated above the scale parameter σ and the 90th percentile of the CDF are also used to characterize the z -error.

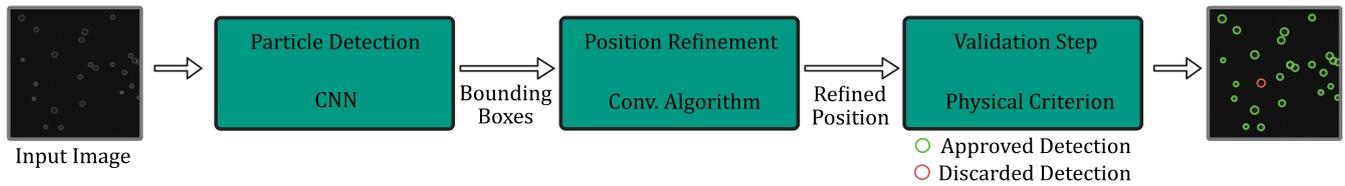


Figure 4. Flowchart of the hybrid approach for particle detection. The input image is fed into a neural network. The neural network detects the particles and outputs a bounding box for each detection. The position of the detected bounding box is subsequently further refined by means of a conventional algorithm. Then the detection is validated in the last step by one or more validation criteria (in this work by an eccentricity limit of the particle image).

The out-of-plane error is indicated in percent of the local z coordinate to avoid dependence on the physical measurement setup and thus accordingly variations of the observed particle image diameter ranges. This error measure is defined from 0 to 1 for the diameter range of the particle images, which is linearly related to the depth of the measurement volume; see equation (2). Note that the neural networks generate a bounding-box for each detection, which comprises information on the center coordinates and box dimensions. The former immediately reveals the in-plane position. The average of box width and height allows a straightforward calculation of the particle image diameter, which is achieved by the conversion of the output formats between the bounding-box format and the circle annotation without any further modifications of the networks output.

2.5. Working principle of the hybrid approach

As already introduced in section 1, this work proposes a novel hybrid approach to particle detection in DPTV. The goal is to combine a neural networks ability to use a broad variety of features to detect particles with the position refinement by a conventional algorithm. The general structure of the hybrid approach is illustrated in figure 4. In the first step, a neural network is used to scan the image and output prediction bounding boxes (detection step). These bounding boxes are then passed on in the second step to a conventional algorithm for position refinement of the bounding boxes (refinement step). In the third step, the refined predictions are then validated by a validation criterion and either passed as the final detection or discarded. The validation step functions as an additional barrier to omit FP and therefore increase the precision. The idea is to add physical knowledge about a particle image to the system, for example to check whether the prediction has reasonable eccentricity for a particle image in DPTV—but also other criteria are feasible. This step is important, since a neural network learns correlations but has no actual physical knowledge of the problem.

The hybrid approach can be seen as the generalized structure for a particle detector and can be employed with any combination of neural network basis and conventional position refinement. The same is true for the validation criteria. The main advantage of the hybrid approach is the decoupling of the detection with the position determination by means of a different approach for each step. Therefore, the individual methods for each step can be heavily specialized for the specific task.

As a result, the necessity in the detection step for high position accuracy is eliminated and the approach can be particularly optimized for detecting particles independent from possible accuracy limitations. Consequently, for the refinement step a broader range of algorithms can be applied that are focused on refining the position of the particle images' bounding box. This offers the possibility of a wide range of combinations of different algorithms and networks. Additionally, specialized approaches can be combined, even though not necessarily performing well individually but working well in combination, thus widening the field of possible combinations further.

3. Empirical evaluation

In this section the two object detection networks faster R-CNN and RetinaNet, with the training status as described in section 2.2, are compared to the Hough transform and the cross-correlation method. First, the methods are tested on various synthetic test data sets to examine the influence of individual effects (e.g. noise and overlaps) separately. Finally, the methods are evaluated in section 3.5 on a set of real experimental data. The insights about benefits and shortcomings of neural networks for particle detection from the empirical evaluation will then motivate the construction and testing of the aforementioned hybrid approach in section 5.

3.1. Synthetic images with randomly distributed particles

To rate the general detection capability for DPTV images, first tests on 200 synthetic images were conducted, similar to figure 2. These images were each comprised of 20–30 randomly distributed particle images each of which having a randomly generated diameter within the aforementioned 15–31 px range. The PR-curve for these test images is shown in figure 5(a). Overall, all detectors achieve precision values of over 99.8%, which shows a generally good detection capability of all detectors. Notable is that—even though achieving a precision of 1—the cross correlation-based method is limited to a recall of 80%, showing that the method is highly reliable when predicting particles but has a miss-rate of 20%. The model function (Hough transform) achieves a recall of 100% but at the cost of lower precision at high recall values. RetinaNet—even though having lower precision than the conventional algorithms—still achieves precision of over 99.9% up to a recall of 99% rendering it a usable

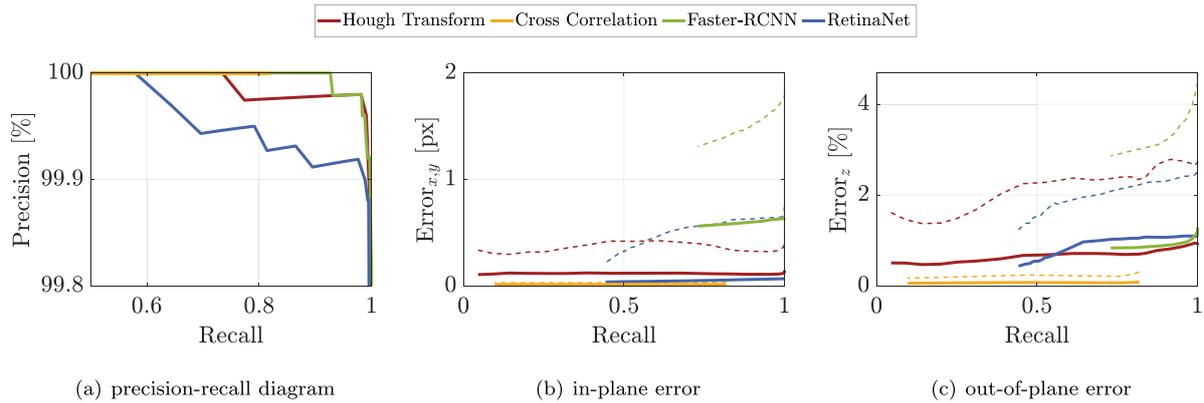


Figure 5. Precision-recall-curve (a), in-plane error (b) and out-of-plane error (c) on synthetic DPTV images with randomly distributed particle images. For the errors plotted over the recall, the scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.

detector. The best detection performance is shown by Faster R-CNN, which reaches recall values of 100% while having higher precision's than the cross correlation or the Hough transform.

While the neural networks demonstrate a strong detection performance, the position accuracy of the neural-network approaches lacks behind the performance of the traditional algorithms. Figures 5(b) and (c) show the absolute in-plane and relative out-of-plane error distributions and its development over a varying recall. It is important to evaluate the error development with respect to the recall, since the rejection criterion is lowered in order to achieve higher recalls, which in turn might also lower the position accuracy of the detector. It is to be expected, however, that a particle image detected only under a lowered rejection criterion—i.e. a particle image more challenging to detect due to deviations of a perfect reference particle image—represents also a bigger challenge to be located with high accuracy. This is shown in figures 5(b) and (c) as the overall trend of an increased position error with increased recall (thus lowered rejection criterion) can be seen. Faster R-CNN has notably higher in-plane and out-of-plane errors, however, still achieves sub-pixel accuracy.

The main drawback of the neural network approaches manifests itself in the significantly broader variation of the error PDF (indicated with the dashed lines in figure 5). Especially RetinaNet achieves lower position error than the Hough transform but suffers from a notably broader error variation for high recalls. However, RetinaNet and faster R-CNN achieve comparable results in the out-of-plane accuracy to the Hough transform, but still can not match the performance of the cross-correlation method. One notable trend for all detectors is that not only the precision decreases with lowering the rejection criterion (to achieve higher recalls), but also the position accuracy decreases. This effect implies also that high recalls come at the cost of not only a decrease in precision but also a decrease in position accuracy due to the inclusion of e.g. distorted and/or spurious particle images.

3.2. Performance at high noise levels

The SNR typically decreases in DPTV when positioning the focal plane further away from the physical particle (to obtain larger particle image diameters) or using smaller physical tracer particles in the fluid. To evaluate the performance at high noise levels 10 random images from section 3.1 were first generated without background noise and subsequently noise was systematically added to achieve the desired SNR value resulting in overall 170 test images. In particular, a data point spacing of 1 has been chosen for $\text{SNR} < 10$ and a coarser spacing of 10 was found sufficient for larger SNR values. For the SNR calculation the definition in analogy to Barnkob and Rossi [18], i.e. $\text{SNR} = \mu_p / \sigma_I$, was used, with μ_p being the mean particle image signal and σ_I being the standard deviation of the noise.

As indicated in figure 6(a), the particle detection decreases in AP for $\text{SNR} \leq 10$ for all detectors. While the cross-correlation method only achieves an AP of $\approx 80\%$ (since the recall is limited to 80%, see section 3.1), it shows the least decrease in detection performance for high noise levels. Furthermore, the neural network based approaches are found to be similarly affected by noise compared to the Hough transform.

While the detection performance of the neural networks is comparable to conventional algorithms for lower noise levels, the position accuracy of the neural networks is significantly more affected by extreme noise compared to the conventional algorithms. This becomes especially obvious from the rapid increase of the variation of the error for $\text{SNR} \leq 10$, demonstrating an increased sensitivity of the machine-learning approaches to increased noise compared to conventional algorithms.

3.3. Overlapping particle images

This section addresses two effects—the influence of the overlap amount between two particle images and the effect of size difference between the overlapping particle images. For the

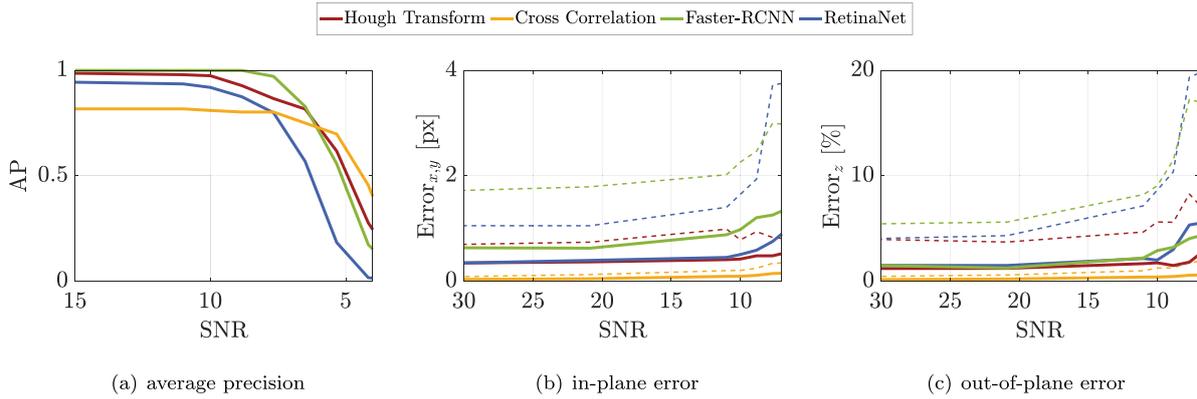


Figure 6. Average precision (a), in-plane error (b) and out-of-plane error (c) over varying SNR. The scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.



Figure 7. Difference between IoU (Jaccard—union area) used for TP definition and the overlap (Szymkiewicz–Simpson—smaller area) definition visualized.

first part, the effect of the size ratio is neglected and the overlap is considered for all ranges of size ratios. Furthermore, this section also focuses on the overlap of two particle images only. The discussion of higher order overlaps (i.e. overlaps with more than two particle images) follows in section 3.4. For a meaningful definition of the overlap a variation of the IoU i.e. the Szymkiewicz–Simpson coefficient [39], was chosen, which rates the overlapping area over the minimum area (therefore over the smaller particle image) instead of the union area (Jaccard overlap), as used for the IoU determinations. The difference between the overlap defined by Jaccard or Szymkiewicz–Simpson is illustrated in figure 7.

This way an overlap value of one means that the smaller particle image is completely inside the larger one, presenting a more meaningful interpretation than the standard IoU to rate the overlap, which takes different values even if the small particle is inside of the large one, depending on their relative sizes. The overlap is mathematically expressed by

$$\text{Overlap} = \frac{\text{area}(A \cap B)}{\min(\text{area}(A), \text{area}(B))}. \quad (9)$$

The test data set used to evaluate the performance on overlap, is comprised of 1000 images containing two particle images each. To vary the amount of overlap and the size ratio of a pair of particle images, their respective in-plane (x, y) and out-of-plane (z) locations were systematically changed over the 1000 images, as visualized in figure 8.

Similar to the influence of increased noise levels (cp figure 6) the detection performance of all detectors decreases with increasing overlap as shown in figure 9. It can be seen

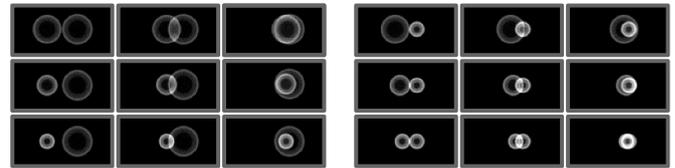


Figure 8. Example images from the test data set to visualize the structure of the overlap and size-ratio determination; top \rightarrow bottom: increasing size ratio, left \rightarrow right: increasing overlap, left group \rightarrow right group: decreasing size of (constant) reference particle-image diameter.

that especially the cross-correlation method is affected even by smaller overlaps (≤ 0.5), thus emphasizing the current problems of overlapping particle images in DPTV. Both the neural networks and the Hough transform show no problems with overlaps smaller than 0.8 but decrease strongly for larger overlaps. Beyond overlaps of 0.8, the neural networks reveal higher APs than the Hough transform, which indicates an advantage over traditional algorithms.

Faster R-CNN has lower position accuracy over the complete range of considered overlaps compared to the two conventional algorithms, which is no surprise due to the generally higher position accuracy of the latter. It is notable, however, that for overlapping particle images, RetinaNet achieves the best position accuracy of the compared algorithms—even better than the cross correlation, which had the lowest position uncertainty in the general case. Especially for higher overlap ratios the position error of conventional algorithms increases significantly to the point where it is comparable with the one of the neural networks.

The second aspect influencing particle image overlaps is the size ratio of the overlapping particle images, since there is a difference between the overlap of two small, two large, or one small and one large particle image. The size ratio of the particle images is, therefore, defined as the ratio of the smaller particle image diameter and the larger one, i.e. $\text{SizeRatio} = d_{\text{small}}/d_{\text{large}}$. To analyze the effect of the size ratio on the detection performance, the achieved recall for each of the 1000 images in the test data set was calculated and located

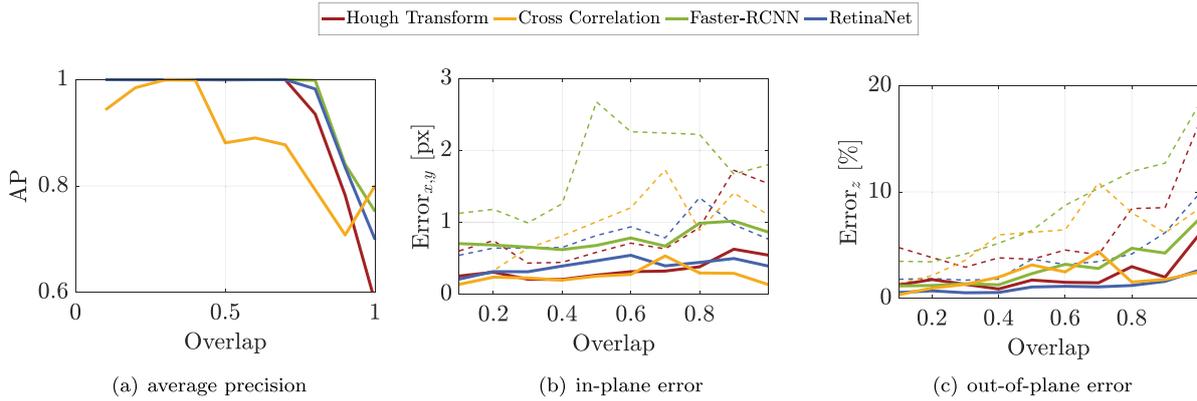


Figure 9. Average precision over the overlap (a), in-plane error (b) and out-of-plane error (c) for varying overlap. The scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.

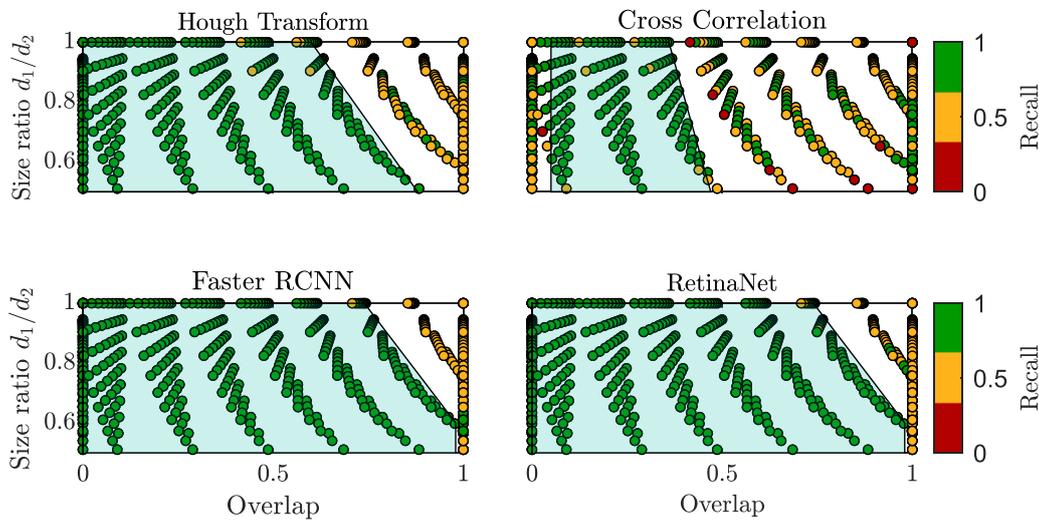


Figure 10. Recall over overlap and size ratio. The recall is color coded in analogy to a traffic light with green (●) denoting the correct detection of both particle images, orange (●) when only one particle image was found and red (●) depicting the detection of neither of both particle images. The overlap and size-ratio regime in which both particle images were detected reliably are highlighted by the green shaded area. The criterion for a reliable detection was a maximum of one orange data point at each overlap-size-ratio position with otherwise only green data points.

in figure 10 based on the overlap and the size-ratio. The recall is color coded to indicate the performance. For better visualization the operating range with respect to the overlap and size-ratio area in which both particle images were detected correctly is highlighted by the green shaded area.

Notably, all algorithms show less detectable overlap values for size-ratios approaching one, which means it becomes more difficult to distinguish overlapping particle images when they are more similar in size. This is to be expected, since two particle images with near 100% overlap and size ratio effectively collapse to a single bright pattern, whereas full overlap of two significantly differently sized particle images still reveals two distinct patterns.

Overall, figure 10 shows that the neural networks can detect overlapping particle images in a much wider range in terms of size ratio and overlap. This is most likely caused by the larger number of features that neural networks use for the detection compared to conventional algorithms. It has to be noted

that the limitation of the neural networks to detect overlaps close to one with size ratios also close to one is likewise a structural problem caused by the non-maximum suppression (NMS). The NMS suppresses duplicate predictions of the network based on an IoU-threshold (here $\text{IoU}_{\text{th}} = 0.8$, standard IoU (Jaccard) not Szymkiewicz–Simpson) that eliminates predictions if they have an IoU with another prediction that is higher than the threshold. This causes correct predictions of overlapping particles to be eliminated by the NMS and was proven by removing the NMS all together (increasing the IoU threshold to 1). In this case the neural networks were able to detect the full overlap and size-ratio spectrum. However, the precision rapidly decreases, since not only correct predictions can pass through but the networks’ multiple predictions of the same object can also pass, causing a large amount of FPs. Therefore, the limitation of neural networks towards overlap is related to the networks structure and cannot be easily solved by just removing the NMS.

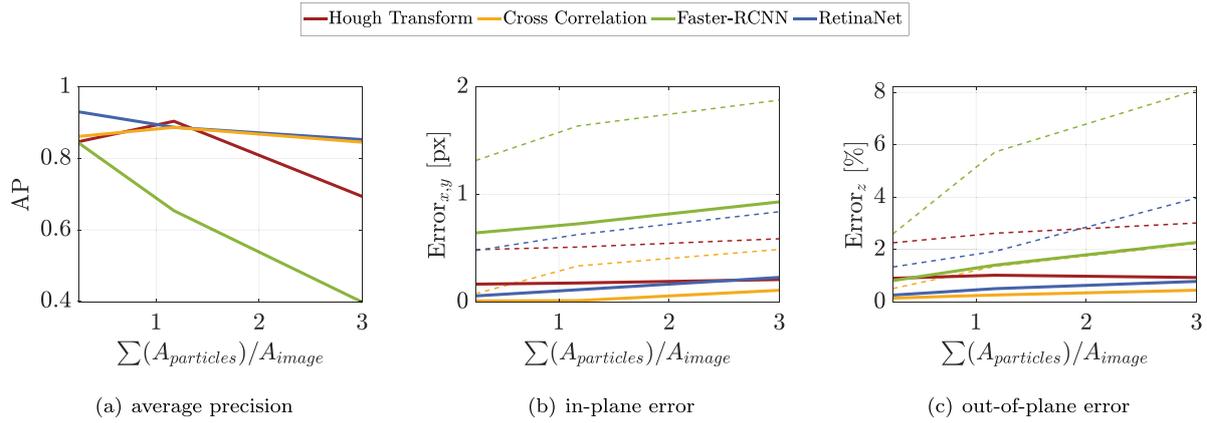


Figure 11. Average precision over the seeding density (a), in-plane error (b) and out-of-plane error (c) for varying seeding density. The scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.

3.4. Influence of seeding density and higher-order overlaps

Since real DPTV applications encounter vast amounts of higher order overlaps (i.e. of more than two particle images, cp. section 3.3) especially when increasing the seeding density, this aspect is addressed separately in this section. Particularly due to the large particle images of DPTV as compared to other volumetric PTV methods, the seeding density needs to be much lower in order to cope with overlap, resulting in e.g. limited information on instantaneous spatial gradients. Therefore, developing algorithms for DPTV, which can deal with more overlaps and consequently with higher seeding densities, is desirable. Conventionally, the seeding density in PIV/PTV experiments is indicated in particle per pixels [1]. However, since the particle size in DPTV significantly changes depending on the z -position, it is more expressive to indicate the seeding density as the ratio of the summed particle-image area over the image area, i.e.

$$N_S = \frac{\sum A_{particles}}{A_{image}}, \quad (10)$$

as proposed by Cierpka *et al* [40]. The test-data set contains 40 images for which every 10 images the seeding density was increased in four steps varying from $N_S = 0.02$ (commonly used seeding density in DPTV) to $N_S = 3$. Consequently, the IoU-threshold defining a TP detection had to be changed accordingly for the calculation of PR. Since, much more frequent overlaps and higher overlap values were present in the image, the IoU-threshold for the definition of a TP detection was risen from 0.5 to 0.8 to avoid a gratification of a shotgun-like detection, i.e. a lucky detection by spamming predictions due to the high density of particle images. Otherwise random predictions could be categorized as TPs by the evaluation algorithm. However, it has to be noted that this makes the PR values only qualitatively but no longer quantitatively comparable to the previous results.

Figure 11 shows the detection results for different seeding densities. Interestingly, the cross-correlation based approach

and RetinaNet achieved good detection performance even for much higher seeding densities than normally used in DPTV, with RetinaNet even slightly outperforming the cross-correlation method with 85% AP at $N_S = 3$. The performance of faster R-CNN, in contrast, rapidly drops with increased seeding density. This is surprising, since one limitation for the higher seeding densities in neural networks is the NMS, which is also present in RetinaNet. However, due to the drastically different performances of faster R-CNN and RetinaNet there has to be another yet unknown mechanism causing further performance issues of faster R-CNN for higher order overlaps. This aspect will be further elaborated in section 4.

For the position accuracy all algorithms show a similar and expected behavior of an increase in position uncertainty with higher seeding densities, while the neural networks retain higher position uncertainty compared to the conventional algorithms. This however is to be expected, since the general position accuracy of the neural networks is weaker, which in turn also leads to the higher position error for two overlapping particle images.

3.5. Performance on images from a real DPTV experiment

While testing on synthetic images allows for the isolation of the desired aspects in DPTV images, in real DPTV applications, other effects such as optical aberrations can not be excluded and have to be dealt with consequently. Therefore, tests were conducted on 30 images containing a sum of 773 hand-labeled particle images from a DPTV experiment [41] as described in section 2.1.

When analyzing the detection performance on real DPTV images all algorithms show a lower precision, see figure 12, which demonstrates the afore-mentioned additional effect of optical aberrations. The conventional algorithms achieve 100% precision only up to 50% recall, which saliently indicates that for a reliable use of these detection algorithms a significant amount of particles in the image has been missed. This

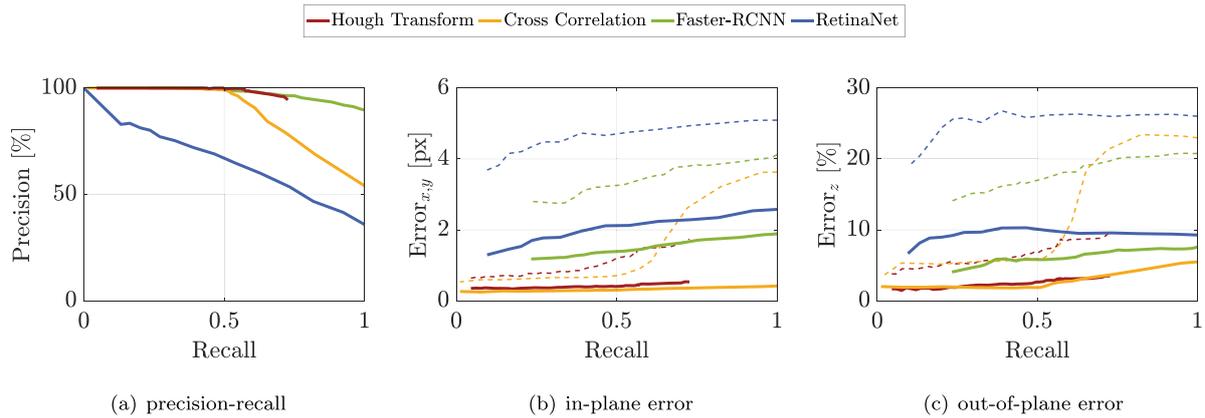


Figure 12. Precision recall curve (a), in-plane error over recall (b) and out-of-plane error over recall (c) for results on images from a real DPTV experiment. The scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.

issue has minor impact for steady-flow experiments, since the lack of detected particles can be compensated with a higher number of recorded images. In case of unsteady flows, in contrast, this issue yields a major disadvantage on the measurement due to the transient flow character.

RetinaNet shows relatively low precision compared to the synthetic test case, indicating that the network did not generalize well from the synthetic training data. Faster R-CNN on the other hand, did generalize well from synthetic training data, showing better detection performance compared to the conventional algorithms.

The poor generalization of RetinaNet to real DPTV images also manifests in the position accuracy, since for real images RetinaNet has higher position uncertainty than faster R-CNN, which was opposite for the synthetic images (cp figures 12 and 5). The conventional algorithms still achieve lower position errors for both in-plane and out-of-plane contributions but only for limited recalls. For higher recalls faster R-CNN has a lower out-of-plane position error than the cross-correlation based method. It appears that poor detection performance correlates to lower position accuracy, which can be explained from the fact that increasing difficulties during particle detection in turn also renders the determination of position and size more challenging. Therefore, in order to achieve higher recalls one has to deal with lower precision and position accuracy, while the quality of the detector determines how much the precision and position accuracy reduce for higher recalls.

4. Findings resulting from the empirical investigations

The application of faster R-CNN and RetinaNet on synthetic images has shown the capability of neural-network based approaches to outperform conventional algorithms in particle detection. This is likely due to the larger number of features used for detection and the significantly larger *a priori* knowledge, which a network gains during training. As such,

the performance of the networks immediately relies on the quality of the underlying training process. That is, the detection performance of neural networks has the potential to increase further, even when already outperforming conventional algorithms on the detection task. Likewise, this can also result in worse performance of a network when insufficiently trained.

The neural-network approaches lack behind on position accuracy compared to conventional algorithms. This problem is of structural nature, since CNNs systematically downsample the image (more precisely the feature maps) resolution while going deeper into the network. The consequently occurring limited image quality accordingly restricts position accuracy. This problem is addressed with the feature pyramid implemented in RetinaNet, which upsamples the feature maps in the network, thus increasing the resolution of the feature maps. Although the achieved position accuracy of RetinaNet was still lower compared to conventional algorithms, the higher position accuracy of RetinaNet compared to faster R-CNN on synthetic images indicates that using feature pyramid networks [28] is a reasonable approach. The feature pyramid is also a possible explanation for the better performance of RetinaNet on the images with higher seeding density, since more complex overlapping particle images can be better distinguished on higher resolution feature maps. Also, most object-detection networks are not developed for sub-pixel accuracy of the bounding box placement, since it is not necessary for most detection tasks. In consequence, highly accurate approaches such as e.g. DPTV image processing as yet has to deal with networks made for detection (and classification) but not for highly precise position accuracy.

Since neural networks excel in detection, they allow to deal with less perfect particle images. Faster R-CNN has shown to maintain a high AP for lower SNR values before it decreased in detection performance. For increased noise the particle-image edges become less distinct from the background, resulting in algorithms like the Hough transform (which relies on the edge gradient) to decrease in performance. In contrast, the overall

optical appearance and the pattern of a bright ring are prevailed much better at high noise levels than the edges. This results in algorithms that rely on more features than the particle image edge to still perform well—even when one feature (e.g. the edge) disappears. Especially faster R-CNN, which is more reliant on texture than on edges compared to RetinaNet [42], should accordingly deal better with higher noise. The neural networks have also demonstrated a robust performance on the detection of overlapping particle images, which again is most likely caused by the usage of a wide variety of features rather than a simple model. However, the networks have a structural limitation caused by the NMS preventing the detection of more extreme overlaps. This limitation is not straight forwardly resolved, since with currently available and applied networks for DPTV it is an essential structural component to avoid duplicate predictions.

Especially for the evaluation of images from real DPTV experiments faster R-CNN has demonstrated superior detection capability and good generalization from synthetic training images, which confirms that training a network with synthetic images can be sufficient for the application on real images (see also [27]). This is an important insight, since training data generation in form of synthetic images is simple and time efficient if generated and labeled with software like e.g. *MicroSIG*, whereas training on real experimental images would face the major problem of ground-truth acquisition for a sufficiently large data set. However, the training of the neural networks for DPTV in the current state is still not optimal. In order to make a network more robust towards noise, optical aberrations and other forms of imperfect particle images, these kinds of particle images should also be included in the training data to allow the network to develop features more robust towards those challenges.

5. Hybrid approach

Overall the neural networks have demonstrated very good detection capability but yield the disadvantage of low position accuracy. Therefore a solution to pair the position accuracy of a conventional algorithm with the detection performance of a neural network would be desirable. As a consequence, the above-introduced new *hybrid approach* is evaluated in this section, which decouples the detection from the position determination (cp section 2.5). By using a neural network for the prediction (detection step) and refining the position with a conventional algorithm (refinement step), the method can maintain the position accuracy of conventional algorithms while utilizing the better detection performance of a neural network.

5.1. Comparison

As consequence from the findings during detection-approach comparison on the images from real DPTV measurements (see section 4), only faster R-CNN was chosen for the hybrid approach. Recall from above that the hybrid approach is

broader and can be employed with any combination of neural-network basis and conventional position refinement. Then the predictions as indicated by the revealed bounding boxes are refined by a simple edge-detection algorithm similar to the one proposed by e.g. Fuchs *et al* [8]. However, instead of an intensity threshold [8] the refinement algorithm used in this work refines the particle image diameter by means of the maximum of the radial intensity. This is achieved by fitting the radial intensity distribution with a polynomial function and determining the maximum of the interpolation to locate the maximum intensity with sub-pixel accuracy. In the last step the detection was validated by measuring the eccentricity of the determined particle image in order to discard false detections for eccentricity values lower than a threshold of 0.5.

The described example version of the hybrid approach was compared to faster R-CNN and the conventional algorithms based on the Hough transform and cross correlation on the images from the real DPTV experiment as described in section 3.5. The resulting performance evaluation of the hybrid approach is shown in figure 13. It can be seen that the hybrid approach reduces the out-of plane position error of the faster R-CNN predictions to the level of the Hough transform, while maintaining the high recall reached by faster R-CNN. For the in-plane error the position refinement resulted in even lower position uncertainties compared to the Hough transform. More importantly, the hybrid approach achieved significantly higher precision than only faster R-CNN (99.8% AP hybrid, 94.4% AP faster R-CNN), where the latter already demonstrated better detection performance than the conventional algorithms. This is caused by the third validation step in the hybrid approach, which introduced additional physical knowledge to the particle detection scheme through elimination of unreasonable detections of the network by validating the eccentricity of the detected particle image.

5.2. Discussion

Overall this simple example of a hybrid approach has shown that this novel concept manages to utilize the quantitative better detection performance of a neural network in combination with high position accuracy of a conventional algorithm. The new approach is furthermore able to improve the detection performance of the neural network by an additional validation step, eliminating most of the FP detections by the network. The most notable characteristic of the hybrid approach however is the decoupling of the particle image detection task from the position determination task. This allows for the usage of neural networks more specialized for detection and eliminates the necessity of high position accuracy in this first step. Therefore, potentially more down sampling deeper into the network and consequently more semantically-rich features should not interfere with the position accuracy of the approach. On the other side, the algorithm used for the refinement step can be more specialized toward sub-pixel accuracy while particle-image detection can be neglected. This offers a lot of room for future improvements in particle detection based on such hybrid approaches.

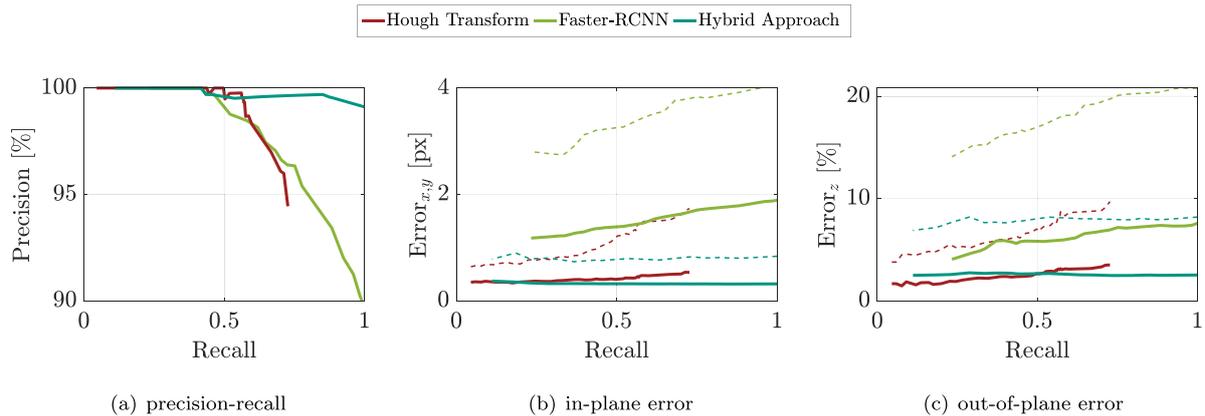


Figure 13. Precision recall curve (a), in-plane error over recall (b) and out-of-plane error over recall (c) for the hybrid approach on images from a real DPTV experiment. The scale parameter σ (peak probability) of the error PDF is displayed as solid lines (—). Furthermore, the 90th percentile of the corresponding CDF is added to the diagrams as dashed lines (- -) to emphasize the distribution of the respective PDFs.

6. Concluding remarks

This paper demonstrated the capability of the hybrid approach to outperform both neural networks and conventional detection algorithms, based on the example of combining faster R-CNN with a simple maximum-intensity detection refinement algorithm. Comparing the neural networks to the Hough transform and a cross correlation algorithm has shown that with sufficient training, neural networks can outperform these benchmark conventional detection algorithms for the task of particle detection in DPTV. The networks also offer the possibility to achieve higher recall values which is beneficial for the measurement of unsteady flows. Such good detection performance is achieved even under compromised circumstances like noise, overlaps or optical aberrations. However, the training of neural networks for DPTV is still largely unexplored and could potentially increase the performance of neural networks toward DPTV even further. Adding more challenging particle images like noisy and highly overlapping particle images to the training data could be the first step toward improved training. Since the used object detection networks do have considerably high capacity in comparison to the limited variation in features of particle images, the networks are likely more prone to over-fitting than under-fitting in training. As a consequence, the inclusion of more complex particle images for a broader training data set is expected to both further prevent over-fitting, as well as increasing the robustness of the network towards detecting challenging particle images. Furthermore a larger variation of the training data increases the generalization capability of the network and therefore, its ability to cope with new and different experimental conditions.

CNNs face a structural limitation regarding particle overlaps due to the NMS, which is not resolvable even with improved training. As such, CNNs without incorporation of NMS are considered particularly candidating for further development. Additionally, neural networks still lack of the desired position accuracy as necessary for DPTV measurements. The lower position accuracy is most likely caused

by the structural problem of the neural network to down-sample feature maps deeper into the network—consequently reducing spatial resolution—and is, therefore, not easily fixed.

A possible solution to this problem is the introduced hybrid approach of the present work: by separating the detection step from the position determination step, a neural network can be used for particle image detection and then the prediction can be passed to a conventional algorithm for a refinement of the location. This hybrid approach combines the excellent detection capability of a neural network with the high position accuracy of a conventional algorithm and therefore effectively circumvents the position-accuracy problem of the neural network. An additional validation step can be added to use a criterion based on physical knowledge (for example checking for a reasonable geometric shape of the particle image) to eliminate false predictions of the neural network, addressing the lack of physical knowledge in the network. In the present study, this additional step demonstrated to improve the precision of the neural network even further.

The hybrid approach offers a lot of new possibilities for combining algorithms, and developing new and more specialized ones for further improvement of DPTV image processing, where the presented proof-of-concept version of a hybrid approach revealed promising results and likewise indicated advanced potential upon further development and/or optimization efforts. As final remark, such hybrid strategies are foreseen to be the basis for future and ongoing developments to advance beyond current limitations of DPTV and APTV image processing, where the present work provides an attempt to contribute to these desired advancements.

Data availability statement

All data that support the findings of this study, including the trained neural networks and supplementary files are uploaded to [KITopen](https://dx.doi.org/10.5445/IR/1000156318) (DOI: [http://dx.doi.org/10.5445/IR/1000156318](https://dx.doi.org/10.5445/IR/1000156318)).

Appendix

Table 1. Training settings for faster R-CNN (ResNet50) and RetinaNet (SSD ResNet50 FPN).

$IoU \geq 0.5$	Faster R-CNN ResNet50	RetinaNet ResNet50 FPN
Optimizer	SGDM	SGDM
Training iterations	50 000	25 000
Minibatch size	1	16
Learning rate α	0.0003	0.04
Learning rate decay	None	Cosine
Warm up period	None	2000
Momentum β	0.9	0.9
L_2 -regularization	None	0.0004
Data augmentation	Horizontal flip	Horizontal flip
Pre-trained	MS-COCO2018	MS-COCO2014

Table 2. Settings for *MicroSIG* used to create the synthetic images.

Parameter	Value	Unit
Magnification	3.5	—
Numerical aperture	0.4	—
Focal length	100	μm
Reflective index medium	1	—
Reflective index lens	1.5	—
Pixel-size	6.5	μm
Pixel-dim x	600	px
Pixel-dim y	600	px
Background-mean	2600	—
Background-noise	85	—
Points per pixel	40	1/px
Number of rays	250	—
Gain	30.2	—
Cyl. focal length	0	μm
z_{min}	-70	μm
z_{depth}	23	μm
Particle diameter	2.5	μm

ORCID iDs

Christian Sax  <https://orcid.org/0000-0003-0749-224X>
 Maximilian Dreisbach  <https://orcid.org/0000-0001-6308-0982>
 Robin Leister  <https://orcid.org/0000-0002-0286-8183>
 Jochen Kriegseis  <https://orcid.org/0000-0002-2737-2539>

References

- [1] Raffel M, Willert C E, Scarano F, Kähler C J, Wereley S T and Kompenhans J 2018 *Particle Image Velocimetry* (Cham: Springer)
- [2] Bryanston-Cross P J, Funes-Gallanzi M, Quan C and Judge T R 1992 Holographic particle image velocimetry (HPIV) *Opt. Laser Technol.* **24** 251–6
- [3] 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
- [4] Elsinga G E, Scarano F, Wieneke B and van Oudheusden B W 2006 Tomographic particle image velocimetry *Exp. Fluids* **41** 933–47
- [5] Schanz D, Gesemann S and Schröder A 2016 Shake-the-box: Lagrangian particle tracking at high particle image densities *Exp. Fluids* **57** 70
- [6] Novara M, Schanz D, Reuther N, Kähler C J and Schröder A 2016 Lagrangian 3D particle tracking in high-speed flows: shake-the-box for multi-pulse systems *Exp. Fluids* **57** 128
- [7] Willert C E and Gharib M 1992 Three-dimensional particle imaging with a single camera *Exp. Fluids* **12** 353–8
- [8] Fuchs T, Hain R and Kähler C J 2016 In situ calibrated defocusing pti for wall-bounded measurement volumes *Meas. Sci. Technol.* **27** 084005
- [9] Kao H P and Verkman A S 1994 Tracking of single fluorescent particles in three dimensions: use of cylindrical optics to encode particle position *Biophys. J.* **67** 1291–300
- [10] Olsen M G and Adrian R J 2000 Out-of-focus effects on particle image visibility and correlation in microscopic particle image velocimetry *Exp. Fluids* **29** S166–74
- [11] 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)*
- [12] 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
- [13] Atherton T and Kerbyson D 1993 Using phase to represent radius in the coherent circle hough transform *IEE Coll. on Hough Transforms* pp 5/1–5/4 (available at: <https://ieeexplore.ieee.org/document/243198>)
- [14] Atherton T J and Kerbyson D J 1993 The coherent circle hough transform *British Machine Vision Conf.* pp 27.1–27.10
- [15] Atherton T and Kerbyson D 1999 Size invariant circle detection *Image Vis. Comput.* **17** 795–803
- [16] Hough P V C 1959 Machine analysis of bubble chamber pictures *Proc., 2nd Int. Conf. on High-Energy Accelerators and Instrumentation* vol C590914 pp 554–8 (available at: <https://s3.cern.ch/inspire-prod-files-5/53d80b0393096ba4afe34f5b65152090>)
- [17] Barnkob R, Kähler C J and Rossi M 2015 General defocusing particle tracking *Lab Chip* **15** 3556–60
- [18] Barnkob R and Rossi M 2020 General defocusing particle tracking: fundamentals and uncertainty assessment *Exp. Fluids* **61** 110
- [19] Rossi M 2020 Synthetic image generator for defocusing and astigmatic PIV/PTV *Meas. Sci. Technol.* **31** 017003
- [20] Lecun Y, Boser B, Denker J, Henderson D, Howard R, Wayne H and Jackel L 1990 Handwritten digit recognition with a backpropagation network *Advances in Neural Information Processing Systems Volume 2 1990* (available at: <https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf>)
- [21] Lecun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition *Proc. IEEE* **86** 2278–324
- [22] 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 *Proc. 13th Int. Symp. on Particle Image Velocimetry*
- [23] 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

- [24] 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
- [25] 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
- [26] 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
- [27] Dreisbach M, Leister R, Probst M, Friederich P, Stroh A and Kriegseis J 2022 Particle detection by means of neural networks and synthetic training data refinement in defocusing particle tracking velocimetry *Meas. Sci. Technol.* **33** 124001
- [28] Lin T Y, Dollar P, Girshick R, He K, Hariharan B and Belongie S 2017 Feature pyramid networks for object detection *Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 936–44
- [29] D’Orazio T, Guaragnella C, Leo M and Distanto A 2004 A new algorithm for ball recognition using circle hough transform and neural classifier *Pattern Recognit.* **37** 393–408
- [30] Rossi M and Barnkob R 2020 A fast and robust algorithm for general defocusing particle tracking *Meas. Sci. Technol.* **32** 014001
- [31] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)
- [32] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* pp 770–8
- [33] Google Brain 2015 Tensorflow: large-scale machine learning on heterogeneous systems *12th USENIX Symp. on Operating Systems Design* (available at: www.tensorflow.org/)
- [34] Lin T Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick C L and Dollár P 2015 Microsoft coco: common objects in context microsoft coco: common objects in context *European Conf. on Computer Vision, 2014*
- [35] Cui Y, Song Y, Sun C, Howard A and Belongie S 2018 Feature pyramid networks for object detection *IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 4109–18
- [36] Bird J J, Kobylarz J, Faria D R, Ekárt A and Ribeiro E P 2020 Cross-domain mlp and cnn transfer learning for biological signal processing: EEG and EMG *IEEE Access* **8** 54789–801
- [37] Bozinovski S 2020 Reminder of the first paper on transfer learning in neural networks *Informatica* **44** 3
- [38] Blair D C 1979 *Journal of the American Society for Information Science* vol 30 2nd edn, ed C J Van Rijsbergen (london: Butterworths) pp 374–5
- [39] Vijaymeena M and Kavitha K 2016 A survey on similarity measures in text mining *Machine Learning and Applications: An Int. Journal* **3** 19–28
- [40] Cierpka C, Rossi M, Segura R and Kähler C J 2011 On the calibration of astigmatism particle tracking velocimetry for microflows *Meas. Sci. Technol.* **22** 015401
- [41] Leister R, Fuchs T and Kriegseis J 2022 Defocusing ptv applied to an open wet clutch—from macro to micro *Proc. of the 20th Int. Symp. on Application of Laser and Imaging Techniques to Fluid Mechanics 2022*
- [42] Rezaei M, Ravanbakhsh E, Namjoo E and Haghghat M 2019 *27th Iranian Conf. on Electrical Engineering (ICEE)* 1589–94