

## Reply to Anonymous Referee #1

July 31, 2025

Dear Reviewer,

Thank you very much for reviewing our manuscript. We are very grateful for the extremely helpful and constructive comments. In the following, we provide point-by-point replies to the points raised in your report. We have marked the original text of the review in blue colour and our response in black colour.

*This manuscript presents CloudTarget, a calibration tool for holographic droplet images, and a CCN-based classification system for distinguishing between particles and artifacts when reconstructing holograms for cloud droplet studies. The authors demonstrate a method they call CloudTarget to quantify droplet detection efficiency, diameter measurement, and position accuracy using a series of chrome photomasks. The methodology is evaluated on several datasets and the performance of the method is evaluated quantitatively. The use of CCN is shown to achieve high accuracy in object classification while enabling faster data processing.*

*This manuscript is very thorough in the reporting of results and it demonstrates high level of technical rigor. Using quantitative metrics and various datasets, the authors systematically compare the performance of their calibration method, CloudTarget, and the CCN used to classify particles and holograms. The paper is well written and organized, however, it can feel dense at some points. As it stands, the manuscript is detailed enough to serve as a good introduction to a beginner calibration methods or holograms of droplet. However, it may be worth considering trimming some content. Overall, I believe this is a useful manuscript to the scientific community and researchers using in-situ holographic data. Therefore, I recommend publication after minor revisions.*

We appreciate the positive assessment of our manuscript, particularly the recognition of its technical rigor and its potential value to the community working with in-situ holographic data. We will address each of the comments in detail, providing point-by-point responses in which we indicate the changes made in the revised manuscript. The response to comment C7 acknowledges and addresses the observation that the manuscript can feel dense at times, mentioned by the reviewer in the general comment. We are thankful for the encouraging remarks and valuable suggestions.

*(C1): Another source of optical artifacts comes from beam quality, which is a result of the quality of the light source, optics used to generate the beam, and any optics that may be used to project an image of the holographic field on a camera. The authors state that artifacts may come from dust and water droplets, however, it seems that the data is processed so that there is background subtraction, so I am unsure how these features might produce artifacts. Can the authors please clarify?*

(A1): We appreciate the reviewer raising this important point.

We have addressed noise sources in the original manuscript in Section 2.1, line 112, where we list potential systematic influences including laser beam inhomogeneities, contaminated optics, and mechanical vibrations. In the revised manuscript we have changed “contaminated optics” to “imperfect contaminated optics” to highlight the fact that of course, as the reviewer correctly pointed out, the optics have an effect themselves, even if they were not contaminated. While our processing pipeline includes static background removal using an advanced median division method based on neighboring holograms, this approach primarily removes persistent features. Noise that appears only in a single or a very small number of holograms, such as those caused by transient dust or water droplets on optics, or sudden beam inhomogeneities, can remain in the data and affect the reconstructions.

Additionally, in Section 2.2, line 157 of the original manuscript, we state that reconstructed planes may contain not only in-focus particles but also diffraction patterns from out-of-focus particles and residual background signals that were not fully removed which we believe is the main reason for our artifacts.

*(C2): Related to the previous point, can the authors comment on why case 1 in Table 2 shows only 22 percent features are droplets? Presumably the other detected features are artifacts?*

(A2): Yes, that is correct, the remaining features are classified as artifacts. The 22% figure in Table 2 refers to the proportion of detected objects that were manually annotated as true droplets in the training dataset. These "objects" that make up training or testing datasets correspond to dark regions identified in the holograms, which can be either real particles or various types of artifacts.

Our CNN-based classifier was trained not only to recognize the appearance of real particles, but also to learn the typical characteristics of noise and artifacts (classification into 2 classes "particle" and "artifact", see line 182 in the original manuscript: "The aim is to achieve a classification into the two classes particle and artifact and therefore the output of the CNN should be a likelihood of the object being in each class."). By including artifacts in the training data, it should improve the accuracy of distinguishing droplets from background or spurious signals in holographic reconstructions.

We have slightly reworded the caption of Table 2 in the revised manuscript to further clarify training and testing datasets include both particles and artifacts.

*(C3): Can the authors discuss how hologram quality impacts the fine-tuning of CNN classifier? To the best of my understanding, the CNN needs to be trained for a specific hologram exposure and image quality. For example, if the beam quality or mean hologram exposure changes from one dataset to another, then the CNN will need to be retrained on a dataset that has similar hologram image quality. Is this true? If this is the case, it would be useful if the authors acknowledge in a direct way that the CNN is highly sensitive to image quality. There is some discussion on this in the manuscript (Section 2.3), but it does not discuss changes in hologram quality.*

(A3): We thank the reviewer for raising this important point. Variations in hologram quality, such as differences in beam uniformity or exposure, can influence the overall signal-to-noise ratio and affect the quality of object detection.

However, the input to the CNN consists of cropped image patches from the reconstruction, which is performed on the filtered and more importantly normalised hologram (normalisation was mentioned in line 134 of the original manuscript). This normalisation helps to reduce the sensitivity of the CNN to absolute intensity differences between datasets/ holograms. In the revised manuscript, we added this information about normalization additionally in line 178, where the CNN classifier input is described, to highlight this point further ("One image is the amplitude and the other the phase, both from the focus planes of the objects determined in the reconstruction of the normalized hologram.").

In our experiments, we observed that the CNN classifier generalizes well across datasets with differing hologram properties. For example, as described in the paper, although the network was trained only on in-situ flight data, it performs well on the laboratory CloudTarget dataset. It also shows good performance on a dataset from the HoloTrack instrument, which includes larger droplets and a different optical setup (see <https://doi.org/10.5194/egusphere-2025-1774>).

Nevertheless, we agree that when applying the classifier to new datasets with substantially different imaging conditions or instrument configurations, it is important to validate and, if needed, improve its performance by finetuning the network with additional training data. We

have updated the manuscript in Section 6.2 and added a bullet point to more clearly emphasize this recommendation. (“While the CNN classifier shows promising generalization from in-situ training data to the CloudTarget dataset, and also to other holographic setups as demonstrated by Thiede et al. 2025, its performance should still be evaluated using established or new verification methods such as CloudTarget before applying it to datasets with significantly different noise characteristics or from different instruments.”)

*(C4): The calibration targets used in the CloudTarget has particles which are deposited on a glass slide. Do the authors think the thickness of the glass slide (and subsequent change in index of refraction and optical path length) change their accuracy to measure droplet location in the z direction?*

(A4): Yes, the presence of any glass surface, such as protective windows in the holographic setup or the glass substrate of the calibration target, can affect the optical path length and thus the measured absolute z-position due to the change in refractive index. In our in-situ cloud applications, we are not concerned with the absolute z-position but rather with relative positioning accuracy and therefore do not correct the effect on optical path length of e.g. protective windows. In the CloudTarget experiments presented in the manuscript, only one photomask was used at a time. The glass of the CloudTarget has a thickness of 2.3 mm and is made of quartz ( $n = 1.46$ ). This would lead to a shift of approximately 1 mm in optical path length if not accounted for, which even for our smallest z-distance of 5cm is only 2%. The absolute z-distances provided throughout the manuscript are primarily used to illustrate trends with depth. In this context, the potential 1 mm optical path deviation is small relative to the measurement uncertainties and does not impact our conclusions. To clarify we have added “Therefore, in the tests described in this paper only a single photomask was used in CloudTarget, which was altered in z-position. The z-positions of the photomask is given as the mean reconstructed distance of the identified particles from the image plane throughout section 5. The change in optical path length due to the refractive index of  $n = 1.46$  of the 2.3 mm thick photomask is approximately 1 mm and neglected here.” in line 321 of the revised manuscript. Further, our evaluation on z-position accuracy focuses on relative z-positioning rather than absolute z-measurements. As stated in the original manuscript (line 690), “We cannot measure the absolute error in z-position but determine how much the z-positions scatter around a perfect 2D plane. Since there is no reason to believe the z-position would have a bias towards over- or underestimation of z, we argue this scatter is a valid measure of accuracy of z-position.” Section 5.4.1 therefore is unaffected by this issue.

*(C5): In Section 5.2.2 a discussion on the diameter dependence of the droplet spatial position is discussed. This effect has been previously reported. For example, see <https://doi.org/10.1088/1361-6501/ab79c6>*

(A5): We thank the reviewer for the reference. While the cited work addresses diameter-dependent spatial effects, to our knowledge it does so in the context of a different imaging modality (not digital in-line holography). In our Section 5.2.2, we specifically discuss and quantitatively assess, how detection challenges arise at the edges of the cross-sectional volume in holographic reconstructions, where parts of a droplet’s diffraction pattern may fall outside the camera’s field of view. This leads to reduced detection efficiency for smaller droplets near the volume boundaries. As such, the underlying mechanism and imaging constraints differ from those discussed in the referenced study.

*(C6): The authors acknowledge that the small glass beads tend to clump together. How are these regions identified and measured? Are they presumed to be artifacts?*

(A6): We thank the reviewer for this observation. The small glass beads were primarily used during early validation tests and were not included in the main results presented in the manuscript. For this reason, we did not include figures or detailed discussions about them, they are only briefly mentioned in Section 3.1.2 as part of the established validation approaches we discussed. In our preliminary tests, we observed that the CNN did classify some clumps of beads as particles.

We have now made that clearer by adding “As at least parts of the clusters are classified as particles in our processing, this makes it increasingly challenging to identify the measured sizes corresponding to individual beads instead of bead clusters.” In line 268 of the revised manuscript. We did not quantify exactly what fraction of the clumps were identified as particles by the CNN.

When measuring monodisperse glass beads, the measured size distribution was clearly multimodal with the first mode corresponding to the bead diameter, and larger modes suggesting the presence of bead aggregates. Manual inspection of the reconstructions for identified particles in the larger modes confirmed that these were often clumps of multiple beads. Although the first peak in the observed distribution aligned reasonably well with the actual bead diameter, the modes were not clearly enough separated which limited our ability further to use these data for fine-tuning the stretch factor in the sizing method.

We want to emphasize however, that the bead aggregates might be less of a problem if the injection method is improved (as pointed out by Reviewer 2, comment 3, and now mentioned in the revised manuscript in line 271) but is not the only limitation of the glass bead method. More important is the inability to compare particle measurements one-to-one to a ground truth as highlighted now more clearly in line 264 of the revised manuscript.

*(C7): As mentioned in the general comments, parts of the paper seem dense. It may be worth considering trimming some content. For example, the Fraunhofer diffraction equations are standard in holography and well-covered in cited references. Another example is the discussion of alternative calibration methods can read more like a part of a review article.*

(A7): We appreciate the reviewer’s comment and fully understand the concern regarding the manuscript’s density. We are aware that the paper is relatively long, however made an effort to keep it concise while maintaining clarity and completeness. We believe the Fraunhofer equation in particular allows insights into why size-, z- and xy- dependent detection efficiency is to be expected. Similarly, we felt that a more detailed discussion of alternative calibration approaches was important to clearly show the novelty and motivation behind *CloudTarget*, and to help contextualize our contribution in relation to existing methods.

Overall, we had a detailed look again into whether the paper could be significantly shortened in the mentioned areas and have removed the detailed description of the reconstructions (lines 134 to 149 in original manuscript) as this is described in detail in Fugal 2009 and a citation seems sufficient. While other steps of our processing chain are also in parts standard procedures, we believe a complete overview of the steps is necessary for reproducibility and transparency.

*(C8): Technical Comments*

- 1. Line 4: consider rephrasing “... with a customized pattern of opaque circles as a verification tool” to something like “... with a customized pattern of opaque circles, serving as a verification tool”*
- 2. Line 20: fix “on board of an aircrafts”*
- 3. Line 122: remove comma “We discovered, that ...”*
- 4. Line 465: there seems to be a grammar error in “from CloudTarget as a reference”*
- 5. Line 604: grammar error in “... and don not see ...”*

(A8): We thank the reviewer for pointing the technical errors out. The issues have been corrected in the revised manuscript.

We provide an additional version of the revised manuscript in which all changes are clearly marked, including those made in response to comments from the other reviewers.

Additionally, lines 141–149, 575, and 597–600 were removed, as they referred to a low-pass filter applied during reconstruction that is not used in our process and was incorrectly mentioned. This has now been clarified in line 137 of the revised manuscript, where Fugal (2009) is cited to describe the reconstruction process and explicitly stated that the method is applied without frequency low-pass filtering. The reference Paliwal 2025 in the list of references was corrected.