The manuscript presents a novel approach for processing data from holographic imagers called HolodecML. HolodecML utilizes GPU hardware to accelerate the reconstruction of the holograms. A neuronal network is trained on patches of the reconstructed images to detect the position of particles. For training, synthetic holograms are created and corrupted to mimic noise that appears in holograms taken by the actual HOLODEC instrument.

The presented utilization of a neuronal network for processing holograms is a very innovative approach, which has great potential to improve the data analysis of holographic images. Additionally, the generation of synthetic holograms with realistic noise is a novelty.

**R2.1:** As these synthetic holograms can be used as ground truth for validating processing approaches because the position and size of the particles are known, the potential of the synthetic holograms could be discussed more prominently.

**A2.1:** We appreciate that you recognize this fact. We mention this in a few places, but it seems reasonable to further emphasize it. To be clear, there are trade-offs in using simulations for this purpose. In particular, overly idealized instrument models are generally not highly performant on real data. In section 2.2 we added the following text to attempt to emphasize this:

“As previously noted, the major benefit to using synthetic holograms for training is that particle positions and size are known and therefore the machine learning solution will not inherit errors or biases from standard processing of actual data. However, the challenge in using synthetic data is that simulations generally fail to fully capture non-ideal aspects of instrument operation, which can impact the effectiveness of the machine learning solution when it is deployed to actual data.”

**R2.2:** In general, the manuscript is well written and the approach is explained in an understandable way. Acknowledging the fact that the authors have the difficult task of presenting a complicated issue at the intersection of computer science and atmospheric science, it still took some time to understand the main message of the manuscript.
In my eyes, the manuscript could be strengthened by focusing on what are the practical advantages of HolodecML in the analysis of holographic data and what is the potential of the new approach. The part of the analysis that is not central to the main points (e.g. analysis with the number of z-slides not equal to 1000) should be moved to the Appendix, whereas the performance comparison should be moved to the main manuscript. The standard approach should be described in more detail to be able to understand the changes and results of the HolodecML approach. For clear reference, the training and test datasets could be given a unique name and their properties could be summarized in a table.

**A2.2:** We feel it is important to establish that the machine learning solution actually works, which is why we have emphasized these metrics and methods in the body of the manuscript. Furthermore, we felt it was important to be clear about the various operational and design tradeoffs in the solution (e.g. the number of z-slides not equal to 1000) because these offer potential performance trades (e.g. speed vs accuracy). This is in contrast to what is often employed in machine learning publications, where the final solution is provided but solution space is rarely described and therefore, not useful to anyone who is not trying to solve the exact same problem.

We uploaded the datasets and a description of them to zenodo. A DOI ([https://doi.org/10.5281/zenodo.6347222](https://doi.org/10.5281/zenodo.6347222)) linking to the datasets is listed in the Acknowledgements section. Additionally, we added Appendix A2 “Data sets” and Table A2 which links the named data files available from the DOI with the named training data splits used in the paper.

**Main comments:**

**R2.3:** The HolodecML approach requires the same reconstruction of planes through wave propagation as the standard approach. Could the GPU implementation of the reconstruction introduced by HolodecML also be implemented in the standard approach? If yes, the performance gains would be of great benefit to the community using the standard approach. Therefore, the performance gains during reconstruction should be separately discussed from the computational cost of the neuronal network for particle detection and the discussion should be moved from the Appendix to the main manuscript.

**A2.3:** Yes, HoloSuite could also leverage this speed up. To clarify this, we moved the content on the GPU acceleration of wave propagation to the section "Estimating $z$ through wave propagation" and added a comment that GPU acceleration could also be applied to HoloSuite. We then moved the discussion on processing speed to the results section.

**R2.4:** I have my doubt that the statement that HolodecML improves particle detection by 20% is based on a fair comparison. The particle detection in the standard approach is normally tuned rather loose to ensure that all particles are detected (i.e., minimizing the number of false negatives). Consequently, a higher number of artifacts that are not real particles (i.e., a higher number of false positives) are detected. The reason behind this is, that particles that are not detected cannot be retrieved at a later stage, but artifacts can be sorted out by a classification algorithm (e.g., by a neuronal network as described in Touloupas et. al, 2020). Therefore, the large false-positive rate and low false negative rate presented in Figure 10 is expected in the standard approach and could be improved by a classification algorithm. How was the particle detection optimized in the standard approach? Were artifacts sorted out by a classification algorithm?

**A2.4:** The reviewer brings up a reasonable point which we struggled with throughout the writing of this manuscript. HoloSuite is modular and has a number of settings that can produce very different results. As a result, it is very difficult to make a complete
comparison to the package. To be clear, the intent of this manuscript is not to declare HolodecML as “better” than HoloSuite, but rather highlight potential approaches that (might very well as likely) be integrated into any processing approach. The comparison presented here represents data from the released CSET dataset. Because this dataset is from a quality controlled field project, we feel that it is a reasonable basis for comparison. We certainly acknowledge that this does not make it the best possible results from the processor, but it does represent a practical result.

We should also highlight the fact that the HoloSuite results also show a higher false negative rate than HolodecML. While this error is smaller than false positives, and, unlike false positives, reduces with increased label confidence, it remains an important and notable metric of performance evaluation, suggesting that the difference between the two methods cannot be entirely made up with a second processing pass on the HoloSuite output.

We have further revised the manuscript to mention that false positives could be filtered out with a second stage classification algorithm in subsection “The standard method and $N_{SH}$ performance on HOLODEC holograms”: “We should note that false positives in the standard method could be improved by adding a second pass classifier to eliminate artifacts from the data. While the data processing considered here did not include this second pass, we still believe it represents a reasonable baseline for comparison since it is part of a released processing run.”

**Minor comments:**

**R2.5:** Figure 1 (a): The interference fringes are almost invisible.

**A2.5:** We have updated all hologram images in the revised manuscript to use a gray-scale for the pixel values (which may range from 0-255 in all images). The HOLODEC example shown in Figure 1 should better show the interference fringes.

**R2.6:** Line 166: What is the RF07 subset?

**A2.6:** In section 2.2, we have clarified that “RF07 refers to ‘Research Flight #7’, which occurred on 19 July 2015 over the Pacific Ocean between Kona Hawaii and Sacramento CA.” For more details on the CSET data set, the reader may refer to Albrecht et al., 2019.

**R2.7:** Line 167: Was HolodecML able to detect the few ice crystals?

**A2.7:** We did not see any while manually labeling the 20 examples from RF07, but we have other datasets where they are present that we are using to extend the modeling approach. We have added the following sentence to the Discussion to note that the current approach could be extended to detect, for example, ice crystals using a multi-categorical segmentation model:

“Finally, the binary prediction task we selected for the segmentation models could be extended to K label types so that other objects such as ice crystals could be identified and characterized”

**R2.8:** Line 201: Is the reference to Figure 3(c)(v) correct?

**A2.8:** We thank the reviewer for pointing out this mistake, the corrected reference is 3(c)(vi).

**R2.9:** Line 203-210: The creation of an independent processing approach is an important motivation for your work and should already be discussed in the introduction section.
A2.9: We have merged lines into the introduction, which is now the second-to-last paragraph:

"Additionally, it was decided that there would be significant benefit in developing HolodecML independently of the current state-of-the-art processing software, referred to here as the `standard method` (and which is discussed below). The motivation for this is twofold. First, by creating an independent processing approach, HolodecML can help identify possible biases and sources of error in the current processing package. Second, this avoided creating a solution where the standard method imposed a ceiling on the processor performance."

The second paragraph in Section 2.4 now begins as:

"In order to develop a processor independent of the standard method, we had to develop a training approach, illustrated in Figure 3(b), that avoided excessive manual labeling (i.e. it is unrealistic to conduct manual labeling of particle position and size over large datasets and the accuracy of such approaches would likely be suspect)."

R2.10: Line 261: In the N = 48648 case the distance between planes is around 3 um. Why did you consider such a large number of planes if you expect limited performance improvement below the in-depth resolution of 57 mm?

A2.10: As the reviewer states, the impact of reconstructing more planes at finer scales than the DOF of the instrument is not expected to produce a significant benefit to the results. However, since we did do the analysis, it seemed worthwhile to include it in the results. Our hope is that this will benefit readers that are not entirely familiar with the concept of DOF. The results of the analysis confirm that adding more planes have diminishing returns for instrument performance.

R2.11: Figure 5: The thresholding in the standard method should detect the particle in all three planes. Why is the mask not visible in all three planes?

A2.11: We thank the reviewer for pointing this out. We have added the following sentence to the caption in Figure 5: "The standard method prediction is the result of a clustering procedure that eliminates the particle in multiple planes."