Point-to-point Responses to Reviewers' Comments

Reviewer #2

We sincerely appreciate the reviewer's critical, physically grounded questions regarding hail impact deformation, shattering, inter-particle collision biases, and limitations of our current manual-placement laboratory calibration protocol. We fully acknowledge that static manual placement eliminates real-world dynamic impact artifacts that occur during natural hailfall, and we will expand error discussion in Section 3.3 Uncertainty Analysis, add new text addressing impact deformation/collision effects, and clearly outline targeted follow-up dynamic drop-test experiments as recommended. Below is structured response, reconciliation of existing results, and concrete manuscript revision plans.

How much would the differing velocities of hailstones hitting the grid surface impact what is imaged? I could see a case where a hailstone hits the grid surface and shatters and/or morphs shape. This would result in slightly inaccurate diameter measurements.

Response: We agree that impact dynamics can influence the post-impact morphology of hydrometeors and therefore potentially affect image-derived size estimation. This effect is physically well recognized in impact-based precipitation measurement systems.

In the case of HailCam, several design aspects mitigate this issue. First, the grid surface is designed as a low-impact interception interface, where hailstones are decelerated prior to imaging rather than undergoing high-energy collision. The segmentation algorithm estimates diameter based on the projected morphological boundary, which remains robust for mildly deformed particles.

Nevertheless, we acknowledge that high-impact events involving large or fast-falling hailstones may introduce deformation or partial fragmentation, which could lead to small biases in reconstructed equivalent diameter. This limitation is now explicitly stated in the revised manuscript as a potential source of measurement uncertainty as follows,

"*Another unaccounted uncertainty source in static manual calibration is morphological distortion caused by terminal-velocity hail impacts on the collection grid. Natural hailstones fall with terminal velocities ranging from ~8 m s* $^{-1}$ *(5 mm hail) to >25 m s* $^{-1}$ *(45 mm hail); high-energy vertical impacts can temporarily flatten oblate hailstones or split brittle ice particles into multiple fragments. The HailCam funnel buffer partially mitigates this by decelerating hydrometeors prior to contact with the imaging grid, and sub-5 mm shattered ice fragments drain through the grid mesh and are excluded from imaging. Even so, partial flattening of intact hailstones widens their projected horizontal silhouette, creating a small positive bias in equivalent diameter retrievals—an error mode entirely absent in our static* *manual placement calibration. This impact deformation bias is not quantified in the present laboratory tests and represents an uncharacterized field uncertainty term.*"

I think it would be advantageous to drop the hailstones during calibration testing as opposed to simply placing them manually. When you manually place them, you remove variability associated with the shattering, breaking, and morphing of hailstones. These are all processes that occur in real life and are likely major sources of error.

Response: We fully agree with the reviewer's criticism and acknowledge that static manual placement constitutes a controlled idealized benchmark rather than a fully realistic reproduction of natural hail impacts. We provide two layers of response: interpretation of existing lab results, plus plans for follow-up dynamic drop testing, plus revised manuscript text to contextualize calibration limitations.

The static manual tests in Section 3.1 were designed to isolate intrinsic optical and deep learning algorithm errors (camera distortion, segmentation contour loss, pixel-to-mm conversion, overlapping particle merging) independent of mechanical impact artifacts. These tests quantify the instrument's inherent optical/AI measurement floor error, which acts as a lower bound on total observational uncertainty. However, we agree that this configuration does not fully represent real hailfall dynamics. We have revised the manuscript to explicitly clarify this limitation in Section 2.3.1,

"*Static manual placement of artificial ice and foam spheres was selected to isolate optical imaging and instance segmentation errors without confounding impact fragmentation or deformation effects. These laboratory results define the minimum achievable measurement uncertainty under idealized particle conditions; total field uncertainty will incorporate additional error contributions from high-speed hail impacts, shape distortion, and fragmentation unquantified in this static framework.*"

We also now explicitly state that dynamic drop-based calibration constitutes an important next step for future system validation, particularly to quantify impact-induced deformation, as follows,

"*Besides, to address the limitations of static manual calibration, a follow-up dynamic drop-test campaign would be helpful using a vertical fall tower capable of reproducing terminal fall velocities for 5–45 mm synthetic ice spheres, in order to develop a quantitative correction model for impact-induced sizing and counting biases to be integrated into the HailCam post-processing pipeline in future firmware updates.*"

Additionally, if images are taken every minute and the hailstones are then funneled out, there is potential for hailstones to collide with and impact one another. The calibration was performed entirely with manual placement, so I am struggling to see how collisions are accounted for in potential sizing errors.

Response: We agree that particle interactions can occur in dense hailfall conditions and may introduce secondary uncertainties in measurement systems relying on collection and imaging.

In the HailCam system, it reduces the likelihood and impact of such interactions that the system does not rely on long-term storage of accumulated hailstones within the imaging region. Even though 60-second imaging duty cycle was used in the filed test, it could be down to 20-second after revise configuration, which would further reduce the chances of particle collision. However, we acknowledge that under high-intensity hailfall conditions, transient overlaps or physical contact between particles may occur. Such events may lead to partial occlusion in images and overestimation or underestimation of individual particle size. We revised manuscript now explicitly includes this as a rare but non-negligible source of uncertainty, particularly during peak hail intensity periods.

"*It also should be noted that over the 60-second accumulation window prior to imaging, newly arriving hailstones collide with particles already resting on the collection grid. These collisions can displace settled hailstones, chip small fragments off particle edges, or press adjacent hailstones into tighter overlapping clusters. Dense overlapping from collision aggregation worsens Mask R-CNN instance merging, exacerbating the undercounting bias observed at high particle densities (Section 3.1.2). The instrument supports flexible configuration of sampling intervals ranging from a maximum of 60 s down to a minimum of 20 s. Shortening the imaging cycle to 20 s drastically cuts the total volume of hailstones accumulated on the tray within each capture window, which markedly reduces the frequency of sequential inter-particle collisions and particle stacking during accumulation.*"

Overall, I think more validation and testing are needed. It seems like the approach was to select the best-case scenario and proceed with the results that appeared most favorable. I would definitely recommend testing with synthetic hailstones being dropped rather than manually placed.

Response: We acknowledge the reviewer's overarching assessment that our current laboratory calibration framework isolates idealized static conditions and excludes key dynamic error modes (impact deformation, fragmentation, sequential inter-particle collision) present in real hailfall. We do not intend to present static lab results as fully representative of all field error sources—this misinterpretation arises from insufficient explicit discussion of calibration limitations in the original manuscript, which we will fully rectify via the following coordinated revisions as shown above, including:

(1) In Section 2.3.1 Laboratory Calibration, add clear text distinguishing static idealized testing (baseline optical/AI error) from untested dynamic impact/collision error sources;

(2) Expand Section 3.3 Uncertainty Quantification to separately discuss impact deformation bias, fragmentation bias, and sequential inter-particle collision bias, explicitly noting none are constrained by existing static calibration;

(3) Add a dedicated future work description outlining terminal-velocity dynamic drop tower testing as the required complementary validation campaign to address the reviewer's concerns;

We emphasize that while static calibration provides rigorous quantification of the instrument's core optical and deep learning segmentation performance, we fully recognize dynamic impact and collision artifacts constitute uncharacterized additional uncertainty terms that require the drop-test validation the reviewer recommends, and we commit to carrying out these experiments to advance the robustness of HailCam measurements in future work.