Response to comment on amt-2023-89 by Anonymous Referee #2

Thanks for your detailed responses and considerable effort to improve the overall quality of the text. It's in excellent shape and I recommend accept with minor technical corrections:

Comment on final figure / line 385 (of original manuscript): Thanks for following up on my poor example, I agree that it's hard to convey this information.

I think the existing plot is suitable given the limitation of not tracking the melting of all individual hailstones.

We highly appreciate, that Referee #2 was available for the second round of review of our manuscript. We agree, that the current representation of the melting process of hailstones in Figure 14 is sufficient. Further, we note that with this minor revision we also make the hail data collection for the given event publicly available (https://doi.org/10.5281/zenodo.10609730) for future analyses.

Figure 2: I can't see the following in the track dots: "The storm location at the sounding time (12 UTC) is marked with the same edge color (magenta)" In (2) I can only see a black rectangle and white cross, no white rectangle and magenta cross.

Unfortunately we forgot to adapt the caption of Figure 2 regarding the black rectangle and white cross. We correct this and add also the magenta edge color marker in subplot (a) again, originally it was there. The resolution of the base map (terrain-background) of Figure 2(b) has been increased, as it was quite pixelated. As raised by the editorial staff, we add copyright statements about the used map material to this Figure: “Map tiles by Stamen Design (stamen.com) and Stadia Maps (stadiamaps.com), under CC BY 4.0. Map data copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org.” In the Figure itself, we additionally give attribution to the Natural Earth data (https://naturalearthdata.com) with an official logo.

Figure 3: HS1 is really hard to spot, took me a full minute, maybe change the color?

Thank you for sharing this experience. Of course we can change the color of HS1 marker. We will use yellow color for HS1 now and add black edge lines around the markers for better visibility.
As raised by the editorial staff, we add copyright statements about the used map material to this Figure: “Map tiles by Stamen Design (stamen.com) and Stadia Maps (stadiamaps.com), under CC BY 4.0. Map data copyrighted by OpenStreetMap contributors and available from https://www.openstreetmap.org.”
To stay compliant with the hail sensor color scheme, the cyan line Figure 12(c) is now yellow.

Figure 5: I can now see why you preferred to use the virtual sensor area instead of center location. Please keep as it is. Thanks for checking this.

Thank you for your comment, we will keep Figure 5(d) as it is.

Response to comment on amt-2023-89 by Anonymous Referee #3

This manuscript presents a case study of using drone-based photogrammetry and deep learning to identify and classify hail size distributions over a soccer pitch in Switzerland. The technique is an advancement of Soderholm et al. (2020) and is a promising way to determine hail size distributions, including the effects of melting, from hail swaths on the ground. The authors compare their results to automatic force-detection hail sensors, radar-based Maximum Expected Severe Hail Size measurements, and a subset of expert evaluations. The manuscript is well-written overall.

We thank Referee #3 for his feedback and comments to further improve our manuscript. Please find our point-by-point answers to these comments below.

Major Comments:

1. I am concerned about the reliability of the small (<6 mm) hail measurements, and I think it would be good for the authors to more directly address and/or plan future follow-ups. These are
    a. ISO 25,600, while not as problematic on modern full-frame camera sensors as in the past, still produces quite a bit of noise. When examining areas on the order of 1-4 pixels, as would be required for hail sizes below 6 mm, areas of noise could very easily be identified as hail. How was mitigation performed?

        So far, no mitigation was performed and a potential solution could be to introduce artificial illumination as discussed in Section 5. There are plans to work on this issue for future data collections.

    b. The authors briefly discuss the impact of motion blur, but for small hail sizes, it could make a larger impact than the authors say. A 1/1000 shutter speed with the drone moving at 1.5 m/s would indicate to me that a single 1.5 mm hailstone could be "smeared" across two neighboring pixels, appearing as a single 3 mm hailstone.

        The drone was flying at a speed of 1 m/s and thus the exact value for motion blur is 0.67 pixels with a GSD of 1.5mm/px. We looked a bit closer to the small detected hail stones and found that there are only 2 classified between 1 mm and 2 mm. Those were not included in our analyses of the HSD, where the smallest bin size was between 3 mm and 6 mm. Thus we correct the total number of hail stones found to 18207 in the text and plot titles.

        In lines 138 and 139 we more clearly specify the motion blur value and expected overestimations:
        "…, which is within one image pixel (0.67 px) in our case and leads in general to small overestimations (≈ 1 mm) of the hail dimensions."

        We agree, that the highest uncertainties are present for the small hailstones. The hail size classes below 1 cm suffer most from the blur effect, which might not be neglectable. Furthermore, wind induced motions of the drone might introduce additional blur. This circumstance is added in Section 2.2:

        "The utilization of a relatively high ISO value, as outlined in Table 1, facilitates operational use even in challenging lighting conditions, maintaining low motion blur

(0.67 px) at a constant flight speed of the drone. Furthermore, wind and gusts can affect the drone's stability, potentially leading to additional image blurring."

The small hail size classes should be further assessed in future analyses. However, hail size below 1 cm does not have a large damage potential and thus we focus on larger size classes. As a concept of proof we currently keep all size classes but in future analyses this effect should be further investigated.

c. To be clear, the values for shutter speed and ISO are reasonable, and the authors discuss the challenges of lack of light. However, more discussion and/or validation at the image collection step in the manuscript would enhance it, which is otherwise not accounted for.

The image collection is one of the biggest challenge in drone based hail observations. This means within extremely short time and under extreme environmental conditions, which cannot be reproduced in a training, the collection must take place in the first try. Thus we designed the collection to be successful within as much illumination ranges as possible. However, various plans to design synthetic experiments for validation of different light conditions should be taken into account for future analyses and are currently planned for our upcoming data collections.

2. I did not see much discussion on how the aspect ratios were determined. I am particularly concerned about the quality of aspect ratio measurements for small hail sizes; I'm a bit perplexed as to how the aspect ratio for small hail is determined given the relatively coarse pixel size versus hail size.

We agree that the quality of the aspect ration might be a concern, as mentioned above regarding the motion blur effecting small hail sizes, future studies should consider a separation of size classes. However, this is out of scope of the current study, but still the results ,e.g. mean axis ratios close to 0.8, are in agreement to other studies like Knight (1986). Despite these quality limitation, there is hardly an alternative to analyze axis ratios of large amounts of real world hailstones. Hence, based on our results, further investigations and an effort to increase the quality should be taken into account for future studies.

Minor Comments:

1. There are several minor grammatical and/or punctuation issues in the manuscript, but I will defer to the copywriting staff to identify and resolve.

Thanks for spotting these grammatical/punctuation issues. We will walk through the text carefully again and the copy-editing stage will hopefully correct the remaining errors.

2. Section 2.1: This section feels too long and not as relevant to the rest of the manuscript.

We note, that this section was already substantially shortened after the first revision and we would like to keep it as it is to share some experiences with the reader. It might also help top plan the data collection process for future field campaigns in particular in complex terrain.

3. Line 152: it would be good to note the temperature in here.

   In principle, this is a good idea, but we lack temperature measurements directly at the hail survey site in Entlebuch. The nearest SMN station, located in Schüpfheim (Table 3), is 5.7 km away. We believe it is more appropriate to address the temperature discussion in Section 4.

   A comment by Referee #4 brings a change. Table 3 has been replaced with a more easily digestible plot, which also displays the start time of the hailfall and the various flight times required to cover the soccer middle circle area.

4. Line 162: How much smaller are the black circles? How does this impact the measurements?

   Thank you for this comment. The black circles against white background appeared approximately 1-2 pixel smaller than the white ones against the black background, likely due to overexposure. This finding does not affect the hail size estimation, as hail is a bright object. We now more specifically write:

   "Due to a slight overexposure in combination with the motion blur, the black circles on white background appeared approximately 1-2 pixels smaller."

5. Line 162: Is there a reason that the overexposure was not corrected for after the fact? Were any highlights in the pictures clipped?

   Because the effect had only a remarkable influence on dark objects against a bright background and not vice versa, we did not consider a correction for overexposure. And no, we did not clip any "highlights". For the analyses part, e.g. the reference objects were masked in black color in order to prevent a false classification of the white circles as hail. In general to ease the image correction for future captures of hail, we strongly recommend to save the RAW images as well, which was unfortunately not the case for the presented event.

6. Line 196: If the additional experts are annotating the same validation and test data as expert A, I'm not sure that these can be described purely as independent comparisons for the ML model.

   The additional human experts (B and C) annotated only the same test dataset, which was not utilized for tuning the model. In this sense, we consider it an independent comparison to assess the ML model results. However, our intention was also to highlight the differences present in the annotations of the experts. It shows another source of uncertainty that is present within the whole process.

7. Line 205: the trademark symbol feels unnecessary

   Thanks, yes this we will remove.

8. Section 3.3: What is the accuracy of the orthophotos? I am concerned that the hail pixels are moving substantially enough that a 1:1 comparison in hail stone size isn't possible.

   As given in Section 2.2, Line 157: the mean GPS error of the first drone flight is 0.34 m. This varies between the flights 1-5 between 0.21 m and 0.5 m. This circumstance makes a simple 1:1 comparison of the hailstone positions impossible. This we clearly state in Lines 323-325. However this does not affect the GSD on the ground which stays (confirmed by the

reference objects) at 1.5mm/px. Thus the alteration of the distributions from the 5 flights within the soccer middle circle can be well attributed to the melting process.

Regarding this we think it is worthwhile to clearly mention this in Section 3.3:
“Due to slight deviations in the derived orthophotos from varying GPS errors (0.21 m to 0.5 m) and the melting process itself,…”

And we add at the end of the paragraph:
“The GSD between the flights stays constant at 1.5 mm/px, as confirmed by the reference objects.”

Response to comment on amt-2023-89 by Anonymous Referee #4

This research represents a somewhat incremental but important step in advancing hail estimates and provides innovative development to address several challenges and make improvements. While building on the work of HailPixel (Soderholm et al., 2020), the authors' key contributions are: (1) a demonstration that the image processing pipeline can be reduced to only applying a region-based convolutional neural network (R-CNN) to an orthomosaic, both simplifying the processing and increasing the accuracy, and (2) an analysis of hailstone melting rates, which is critical to accurately understand any post-event hail observations, using successive flights. Furthermore, the data collection procedures presented continue to refine strategies for the successful interception and observation of hail events, which is non-trivial. There are some details of procedures and implementation that are not fully optimized but seem within reason for initial experimental purposes. Overall, the greatest limitation of this work is the very limited data set (a single event), which makes it difficult to understand the broad applicability of some specifics (for example, the R-CNN model as-trained or loss of accuracy due to lighting conditions). However, it represents valuable proof-of-concept with novel approaches and would serve as a steppingstone for future work. The techniques applied lend themselves easily to such future development and expanded data collection, relying only on commercial-off-the-shelf equipment and consumer-grade computing equipment.

We thank Referee #4 for his positive evaluation and comments to further improve our manuscript. Your observation is indeed right. We perceive this contribution as a continuation following the HailPixel research by Soderholm et al. (2020). Subsequent advancements are anticipated to build upon this foundation. And with more captured hail events, the techniques to identify hailstones and calculate the HSD can be finetuned, and hopefully more generalized at some point.

Comments/Questions/Suggestions:

1. Line 56 "lake" -> "Lake"

   Typo has been corrected on Line 56.

2. Line 62 "for the area within a distance of less than 1 km from the survey area" -> "for the area within 1 km of the survey area"

   We simplified the sentence as suggested.

3. Lines 69-70 "a high resolution" is ambiguous, what is the resolution? (It is given on Line 126.)

   Similarly, Line 53 "giving a ground sampling distance (GSD) of 1.5 mm px−1" From what altitude of flight? (It is given on Line 137.)

   It would be nice to have all of this information presented in a single statement or at least the same section (e.g. something like "A ground sampling distance (GSD) of 1.5 mm px−1 was achieved flying at an altitude of 12m with a 45 MP camera."), similar to how it appears in your abstract. Perhaps even a flight characteristics table would make referencing this comparatively in future work easier. This is not a critical point as all the information is presented, but as a matter of preference could be easier to consume.

We agree that it is useful for the reader to have these information altogether in one sentence. In Lines 69-70 we now specify:

"Following those suggestions, we achieved a ground sampling distance (GSD) of 1.5 mm/px by flying at an altitude of 12 m with a 45 megapixel full frame camera system. For the detailed flight and system characteristics see also Table 1."

In Section 2.2 we incorporated a new table (Table 1) to summarize the drone, camera and flight characteristics of the main mission ("Specifications of the drone, camera system and flight characteristics."). Thus we shortened the belonging text parts a bit by linking to the information in Table 1.:

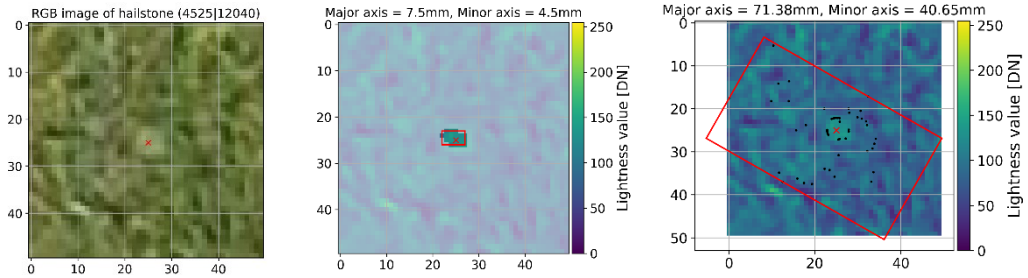
"Table 1 summarizes the detailed drone, camera system and flight characteristics of the hail photogrammetry mission."

4. Line 91 "better", Is it possible to present a quantified comparison for this?

Sorry for not being very precise in our formulation. We now specifically say:

"With the two-stage approach, the edge detection did not work reliable in particular for small hail stones, because the lightness gradient between these small hail stones and the background was insufficient. Here we therefore focus on the one-stage approach."

A more quantitative and in depth analyses of the performance of edge-detection algorithms for small hailstone sizes will for the moment be left to future research. This in general is also connected to image quality, noise and motion blur. Here we just show a simple example of a small hailstone (left image), where our edge-detection (right image) obviously failed:



The segmentation mask from the R-CNN model (middle image) is much closer to the true stone shape.

5. Section 2.2. Did you have any guidelines for acceptable flight conditions? E.g. Maximum wind speed/gusts, etc. It is clear you were trying to get off the ground as soon as possible after an event, but given the criticality of timing, it would be helpful to know if there were any additional limitations.

Although the Matrice 300 reaches IP45 water protection, there is no product warranty to cover water damages. It is the risk of the operator when flying in wet conditions. We followed our defined rule to not fly the drone if the rain rate exceeds 4-5 mm/h. Safe operation is guaranteed up to a wind and gust speed of 10 - 15 m/s. These values of course depend on the drone system in use.

In Section 2.2 we suggest to add the following sentence: “Although it is critical to get off the ground as soon as possible after the hailfall, environmental conditions like rain rate, wind and gust speed should be carefully watched out in order to stay within the permitted operation conditions of the drone model.”

6. Line 154 “ISO-25600” This is very high and likely introduces a fair bit of noise. This is explained later on Line 428, but I think it would be useful to include the reasoning for such a high ISO in the earlier section. Could results be improved by running a slightly slower shutter and lower ISO? Is wind playing a role in image blur in addition to forward flight? These may be topics for future work, but it would be nice to see them acknowledged if applicable.

   We think that much more field tests are needed to find the optimal settings for the camera dependent on the available light conditions. Sensitivity tests need to be performed and could be another step towards future improvements of the method. There is also the possibility to install a light source next to the camera with a dual-gimbal mounting device, which could be very useful to reduce the ISO by keeping a fast shutter speed. We discuss the potential of a light source and state that the image quality can be improved by reducing the ISO in Section 5. We also included a point in the new list of ideas:

   - Fine tuning of hardware settings and flight characteristics for optimal image quality in conjunction with an acceptable motion blur.

   The reasoning for the high ISO is added now in the earlier Section 2.2:
   “The usage of a quite high ISO (see Table 1) value is explained by simplifying the operational usage also in difficult light conditions, while keeping the motion blur low (0.67 px) at a constant flight speed of the drone. …”

   Yes, wind and especially gusts can reduce the stability of the drone flight and thus have the potential to add additional image blur. But it is really difficult to quantify the contribution of wind to the blur effect and is currently out of scope for this work.

   To the paragraph above (within Section 2.2) we add: “In addition, wind and gusts may influence the stability of the drone movement, which could further blur the image.”

   Regarding image quality it would be best to program the drone to stop at each capture point to remove the image blur from forward motion. However the amount of time then would largely increase and lower the size of the area that can be mapped. There are tradeoffs one has to decide on.

7. Line 157 GPS Error: Are you using RTK correction? The error value suggests not, even though your drone supports it, and you call this out explicitly (Line 127). There’s obviously a substantial challenge in deploying an RTK base station and establishing a usable dilution of precision in the timeframes you require. However, it would be worth mentioning these limitations and maybe potential alternatives, such as NTRIP (Networked Transport of RTCM via Internet Protocol) services, if available, especially given the discussion in Section 3.3 of lacking positional consistency without ground references such as the soccer center circle.

   Thank you very much for this comment. You are right we called the RTK feature out, but actually did not operate the drone with an RTK base station. As you say, it introduces a lot of additional efforts. We will clarify this in the text also with regard to the positional inconsistency between the flights.

In line 127 we now do not mention the RTK feature and clarify thereafter:

"The drone was not operated with enabled RTK (Real Time Kinematic) feature. This would require the installation of a RTK base station module. The advantage would be an increase in positional accuracy of the drone from the order of few decimeters to centimeters. Another potential option would be to use the NTRIP (Networked Transport of RTCM (Radio Technical Commission for Maritime Services) via Internet Protocol) protocol. This protocol facilitates the transmission of correction data over the internet. It enables real-time positioning and precise navigation by delivering accurate correction data to GPS receivers."

In Section 3.3 , also with respect to a comment by Referee #3, we specify the GPS errors more detailed now: "…from varying GPS errors (0.21 m to 0.5 m) …"

We note, that we unified the appearance of the expressions "soccer middle circle" and "soccer center circle" to "soccer middle circle" in the whole text.

8. Line 297 "equal axis lengths". From your bin definitions, this actually represents aspect ratios >0.9, not necessarily exactly equal.

   Thanks for spotting this, here we change the formulation now to: "The projected hail aspect ratios indicate that the majority of hailstones are rather spherical with axis ratios greater 0.9 (Fig. 11(b). 75 % of the hailstones have projected aspect ratios higher than 0.75."

   Lines 403-404: Consequently we changed here to: "The median hailstone size was 9 mm and the majority of hailstones were rather spherical with axis ratios greater 0.9."

9. Figure 5(d) doesn't make sense until you get to Line 300. I understand why it makes sense to have with (a)-(c) as a single figure, but it may be worth noting the section it applies to in the caption.

   Now we added in the caption of Figure 5 that plot (d) belongs to Section 3.2.

10. Lines 306-309 might be more easily digested as a table, but that's more a preference.

    Here we would like to stay for now with the 3 sentences, unless it is more strictly recommended to use a table instead.

11. Section 4 – First paragraph could maybe go in intro, feels a little out of place here, but again more of a preference.

    In regard of the discussion of different techniques to identify hail stones in the subsequent paragraphs, we believe the first paragraph here does fit quite well, as the transparency of the hail stones is connected to the performance of the different techniques.

    We found that the last sentence here ("Thus, for hailstones with high transparency the approach used here might not work.") is sort of confusing and we now write for a better transition instead: "The effectiveness of various methods used to detect hailstones is influenced, in part, by the transparency of the ice."

12. Line 367 “as published in e.g. Knight (1986); Shedd et al. (2021)” -> “as published in Knight…”

    Line 367: This has been adopted, “e.g.” is deleted now.

13. Table 3 – Would it be possible to make this as a time series plot, T on left axis, RH on right (or similar)? It is more difficult to pick out the trends looking at a table. Noting the flight times as vertical lines or highlighted sections would further help in understanding the overall timeline of events.

    Thank you for this suggestion and we think it is a good idea to put the temperature, information in a simple time series plot together with time information of the start of the hailfall (vertical black line) and drone flights (grey shaded vertical bars) to capture the soccer middle circle (as given in Table 2). Here is the version we suggest and we put into the manuscript:

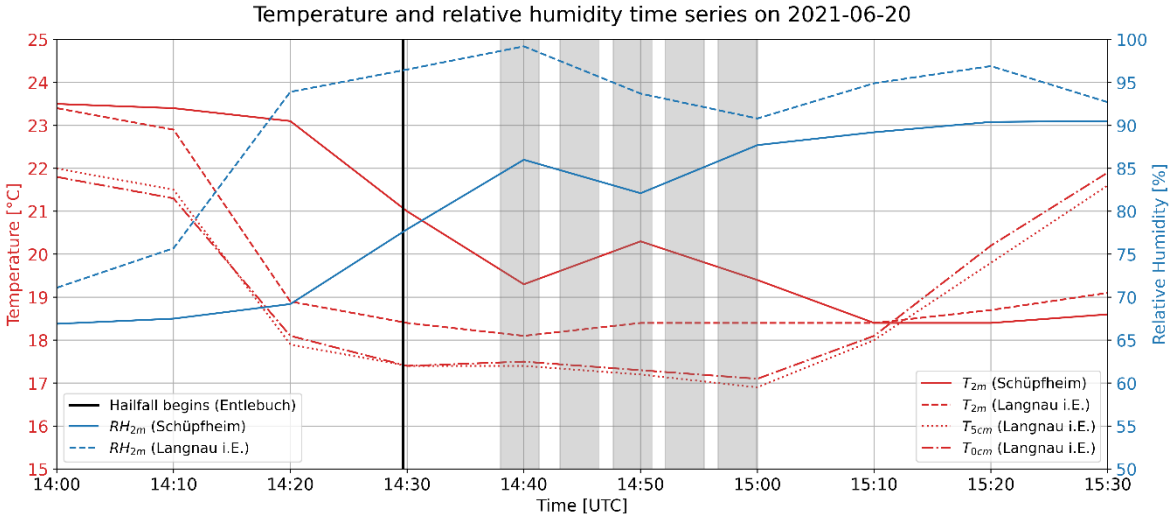


    Table 3 is then obsolete (all information is in the plot) and will be deleted.

14. Line 385 “what might effect” -> “which might affect”

    Line 385: We changed the wording from “what might effect” to “which might affect”.

Overall largest concern: Is this repeatable and generalizable? Your results are based on a single event used for training, validation and testing. Very interesting work though and sets the stage for future research that can begin to fine tune and hopefully more extensively validate these types of analyses.

As we describe in Table 4, the operational application of drone-based photogrammetry of hail is challenging and thus requires time and resources to build larger data sets incorporating several events. For looking into the broad applicability more data from hail events are needed. We hope that in a future, more researcher successfully execute such drone flights in the field and the community puts an effort in building up a database of aerial images. Such a database could help to further develop this approach and build more generalized models.

Other suggestions for potential future research:

(No expectation of these for this publication but curiosities that may be of interest to the authors.)

- Integration of thermal imagery. Even with low resolution, the integrated pixel values could provide useful information. By using the surface temp in areas with high probability and confidence of not having hail present as a background, you could use the differentials of other pixels to help include or exclude potential hailstones in conjunction with RGB techniques.

- Utilizing SfM result and applying R-CNN directly to mesh or point cloud rather than 2D orthomosaic. This would obviously require more computing power, but it would be interesting to see how it changes performance in 2D visually challenging environments (like taller grass).

Thank you very much for these two additional suggestions for future work and possible improvements of the method. We decided to take those ideas and include them in the outlook of the paper in Section 5:
Other ideas to test and potentially improve the techniques in the future are:

- Integration of thermal imagery to help exclude or include potential hailstones alongside the RGB image processing.
- Usage of SfM (Structure from Motion) results and application of Mask R-CNN directly to mesh or point clouds instead to the RGB orthophotos.