Review of “Drone-based photogrammetry combined with deep-learning to estimate hail size distributions and melting of hail on the ground”

Summary:
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.

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?

    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.

    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.

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.

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.
2. Section 2.1: This section feels too long and not as relevant to the rest of the manuscript.
3. Line 152: it would be good to note the temperature in here.

4. Line 162: How much smaller are the black circles? How does this impact the measurements?
5. Line 162: Is there a reason that the overexposure was not corrected for after the fact? Were any highlights in the pictures clipped?
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.
7. Line 205: the trademark symbol feels unnecessary
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.