Response to Reviewer 1

"Automated Analysis of Ripple-Scale Gravity Wave Structures in the Mesosphere Using Convolutional Neural Networks"

Jiahui Hu et al.

We thank reviewer for the careful and constructive evaluation. The comments led us to expand the algorithm description, clarify the physical interpretation of ripple-scale structures, strengthen the quantitative validation, and improve the figures and discussion. The reviewer's original comments are reproduced below in regular type. Our responses are shown in *slanted type*.

Major points

Reviewer 1:

> This manuscript is intended for publication in AMT. The focus of AMT is on new measurements, methods, algorithms, etc. The manuscript's innovative aspect is its use of a machine learning approach to derive small-scale wave structures (ripples) from OH airglow measurements. Consequently, the focus is on the algorithm. It would be helpful for readers and other airglow scientists operating similar instruments who might want to apply this or a similar approach if the algorithm were described in more detail and in an easier-to-follow way.

Response:

> *We thank the reviewer for this helpful suggestion. We agree that, because the primary contribution of this study is methodological, the algorithm should be described in sufficient detail for readers to understand and reproduce the processing workflow.*
>
> *We therefore substantially expanded the Methodology section to describe the complete detection pipeline. The revised manuscript now explains the construction of signed two-minute image differences, the spatial calculation of the median and median absolute deviation (MAD), and the normalization and clipping applied before patch extraction and full-image inference.*
>
> *We also expanded the description of the SE-CNN architecture. The revised text defines feature maps, channels, convolutional filters, ReLU activation, and the dimensions C, H, and W, and Table 1 provides the layer-by-layer network architecture. The squeeze-and-excitation operation is additionally described mathematically through the squeeze, excitation, and scaling steps.*
>
> *For full-image inference, overlapping 41 × 41-pixel windows are evaluated with a stride of 5 pixels. The positive-class softmax output is treated as an uncalibrated model score rather than a numerical probability. Candidate regions are generated using a model-score threshold greater than 0.99, followed by spatial merging and component filtering based on*

pre-clipping area and eccentricity. Analysis is restricted to the centered 125-pixel-radius region of interest. The provisional temporal analysis then links retained components in consecutive frames using the stated temporal-separation and centroid-displacement criteria.

We also added explicit definitions of frame-level spatial recovery, maximum-cardinality one-to-one object matching, and the blinded disagreement-patch audit. These revisions distinguish patch-level classifier performance from full-image validation and clarify the limitations of the selected validation design.

Reviewer 1:

Even though the focus is on the algorithm and the results do not provide more information than has already been extracted manually, the discussion could be improved. There is no comparison with other publications, and statements are not supported by citations. Consequently, readers may get the impression that the discussion was prepared rather hastily. A similar impression is given by the results section. Figures 4 and 5 are described but not referenced in the text. Figure 6 is not described at all (or I was not able to attribute the description to the figure), and for Figure 7 it seems that either a different figure is described or important information is missing.

Response:

We thank the reviewer for this assessment. We agree that the original Results and Discussion did not sufficiently place the automated method and its results in the context of previous studies, and that the figure descriptions were not sufficiently integrated with the text.

We have revised the Results section so that the principal figures are explicitly introduced and interpreted in the surrounding text. Figure 4 is now discussed as a comparison of provisionally linked reference and automated tracks within the selected validation cohort, while Figure 5 is discussed as a seasonal comparison of frame-, object-, and track-level counts. The revised text also explicitly states that these quantities are conditional on the selected manually positive validation cohort and are not archive-wide, exposure-normalized occurrence rates.

The Discussion has also been expanded to compare the present observations and methodology with previous studies of ripple-like and small-scale wave structures. We now discuss the differences in seasonal behavior, local-time and directional variability, duration, and detection methodology, while emphasizing that numerical results from different studies are not directly interchangeable because of differences in site, instrument, observing period, preprocessing, labels, thresholds, occurrence denominators, and evaluation units.

We also revised the discussion of event duration. Because the manual catalog does not contain independently defined event start and end times, comparable manual lifetime statistics cannot be established. We therefore no longer interpret the former manual lifetime distribution as observational duration data and instead describe the automated durations as products of the provisional temporal-linking procedure.

Reviewer 1:

Finally, ripple structures have long been interpreted as part of the gravity-wave breaking process. Li et al. (2017) demonstrated that these small-scale wave structures can also be secondary waves. These authors used not only airglow images but also additional measurements. Since such measurements are unavailable at most airglow observation sites, it is not possible, as far as I am aware, to discriminate between instability features and secondary gravity waves based on horizontal wavelength alone. The manuscript neglects the possibility that small-scale waves may be secondary waves. This should be mentioned.

Response:

We thank the reviewer for this important observation. We agree that the original manuscript placed too much emphasis on gravity-wave breaking as the physical interpretation of the detected structures.

We have revised the manuscript to make clear that the SE-CNN identifies ripple-like image morphology rather than a unique physical generation mechanism. As discussed by Li et al. (2017), small-scale structures with similar airglow morphology may also represent secondary gravity waves. Because the present analysis uses OH airglow imagery alone and does not include coincident wind and temperature measurements, the automated detections cannot be uniquely classified as instability features, gravity-wave breaking, or secondary waves.

We therefore use "ripple-like structure" as a morphological term throughout the revised manuscript and explicitly state that distinguishing among possible generation mechanisms requires complementary observations beyond the scope of the present study.

Detailed comments: General

Reviewer 1:

Line numbering would have facilitated the review.

Response:

We thank the reviewer for this suggestion. Continuous line numbers have been added throughout the revised manuscript to facilitate review and precise cross-referencing.

Detailed comments: Introduction—Formal

Reviewer 1:

Parentheses around citations are occasionally missing (e.g., page 2, end of first paragraph: Fritts and Alexander (2003)).

Response:

We agree. Citation formatting has been revised so that narrative citations are used when the author name is grammatically part of the sentence, whereas parenthetical citations are used when the reference supports a statement without forming part of the sentence. We also performed a citation-format consistency check throughout the revised manuscript.

Reviewer 1:

Abbreviations are not always defined (e.g., YOLO, ASI, CCD).

Response:

We agree. Abbreviations are now expanded at first use and used consistently thereafter. In particular, the revised manuscript defines charge-coupled device (CCD) and You Only Look Once (YOLO) at first use. We also use "all-sky imager" or "all-sky airglow imaging" explicitly where this improves readability rather than introducing an unnecessary abbreviation.

Detailed comments: Introduction—Content

Reviewer 1:

As mentioned above, ripples are not necessarily part of the gravity-wave breaking process; they can also be secondary gravity waves, as demonstrated by Li et al. (2017). Please clarify this.

Response:

We agree. The Introduction has been revised to state explicitly that ripple-like structures are a morphological category and cannot be assigned uniquely to gravity-wave breaking or instability from the airglow morphology alone. We now note that similar structures may also represent secondary gravity waves, following Li et al. (2017), and that distinguishing these mechanisms requires additional observations.

Reviewer 1:

There are relatively few studies using machine learning to extract small-scale waves from airglow images. To my knowledge, there are three such studies:

- Lai et al. (2019),
- Prasad et al. (2025),
- Wüst et al. (2026).

It would be worthwhile to mention all three studies and compare your approach (perhaps in the Introduction) and results (perhaps in the Discussion) with theirs.

Response:

We agree with the reviewer. The Introduction now discusses Lai et al. (2019), Prasad et al. (2025), and Wüst et al. (2026) as related machine-learning approaches. We clarify that Lai et al. focused on gravity-wave extraction, whereas Prasad et al. and Wüst et al. used object-detection approaches for localized small-scale structures. We explain that the present method differs by classifying 41 × 41-pixel patches and reconstructing a full-image score field using overlapping sliding windows. The Discussion now compares these approaches while noting that differences in instruments, labels, preprocessing, and evaluation units prevent direct numerical comparison.

Detailed comments: Methodology—Formal

Reviewer 1:

Figures 3 and parts of Figure 4 are difficult to interpret in print. Increasing line thickness and symbol size would improve readability. The white line on a black-and-white background should also be reconsidered.

Response:

Implemented. Revised Figure 3 uses thicker gray, bluish-green, and blue score contours with larger orange manual markers and blue model markers. Revised Figure 4 uses larger markers and two stacked panels. White contours and red/green-only distinctions have been removed.

Detailed comments: Methodology—Content

Reviewer 1:

Section 2.1: Taylor and Hill (1991) did not investigate the OH airglow layer height. More appropriate references include:

- Baker and Stair (1988),
- Shepherd et al. (2006),
- von Savigny (2015),
- Wüst et al. (2020).

Response:

We agree with the reviewer. Taylor and Hill (1991) was not an appropriate reference for the nominal OH-emission-layer altitude. We therefore replaced it with references directly addressing OH-layer altitude and its variability, including Baker and Stair Jr. (1988), von Savigny et al. (2015), García-Comas et al. (2017), and Wüst et al. (2020).

We also revised the wording to describe approximately 87 km as the nominal OH emission peak altitude while explicitly noting that the layer altitude varies temporally and geographically.

Reviewer 1:

The abstract states that time-difference images are used, but this is not explained in Section 2.1. Please clarify and discuss which waves are enhanced or suppressed by this procedure.

Response:

We thank the reviewer for this helpful comment. We have expanded Sect. 2.1 to explain both the construction of the time-difference images and the associated selection effects.

For each target image, the immediately preceding valid image acquired approximately 2 min earlier is identified, and a signed difference is calculated. The resulting operation

acts as a temporal high-pass filter whose response depends on the image cadence and the apparent motion and temporal evolution of the observed structures. Stationary and slowly varying features are suppressed, whereas structures that change intensity or translate between consecutive frames are enhanced.

We also note that slowly evolving ripple-like structures may therefore be attenuated, while translating structures may generate paired positive and negative residuals corresponding to their apparent movement between frames. Rapid motion may additionally reduce the spatial coherence of the residual. We now explicitly state that time differencing modifies the effective selection function of the detector.

Reviewer 1:

Equation (1): Is the median computed over space, time, or both?

Response:

We thank the reviewer for identifying this lack of clarity. The median and median absolute deviation (MAD) are calculated over the spatial dimensions of each complete signed time-difference image. Specifically, after computing $(D_t(x, y) = I_t(x, y) - I_{t-\Delta t}(x, y))$*, the median is calculated over all image pixels, followed by the MAD relative to that spatial median. This calculation is performed independently for each time-difference image. No temporal median or temporal averaging across multiple frames is used. The revised equations and text in Sect. 2.1 now make the spatial nature of this normalization explicit.*

Reviewer 1:

Please specify the all-sky camera field-of-view diameter and maximum zenith angle to help readers estimate the physical dimensions corresponding to the 41×41 pixel patches.

Response:

We thank the reviewer for this suggestion. The revised Sect. 2.1 now specifies that the all-sky imager has an approximately 180° field of view, corresponding to a nominal maximum zenith angle of about 90°.

We also clarify that the 41×41*-pixel patch does not correspond to a constant physical horizontal area because of the fish-eye projection. The horizontal distance represented by a pixel, and therefore the physical dimensions represented by the patch, vary with zenith angle. The present validation is restricted to a centered 125-pixel-radius region of interest, and the CNN operates in image coordinates rather than assuming a constant physical scale across the field.*

Reviewer 1:

Section 2.2 requires a substantially clearer explanation of the neural network for readers unfamiliar with machine learning. For example:

- Explain "channel" and "feature."
- Describe the filters used.

- Define ReLU.
- Explain the variables C, H, and W.
- Provide references introducing CNNs and justifying design choices.

Response:

We thank the reviewer for this helpful suggestion. We expanded Sect. 2.2 to make the neural-network description accessible to readers without a machine-learning background.

The revised manuscript now explains that a feature map is a learned spatial response produced by a convolutional filter, defines a channel as one such feature-map dimension, describes the role of convolutional filters in detecting local spatial patterns, and defines the ReLU activation as ($\mathrm{ReLU}(x) = \max(0, x)$). *We also define C, H, and W as the number of channels, feature-map height, and feature-map width, respectively.*

In addition, Table 1 provides the complete layer-by-layer architecture, and the SE operation is described mathematically to clarify how channel-wise recalibration is performed.

Reviewer 1:

Figure 2: The term "vanilla CNN" appears before being explained (and contains a spelling error). Additionally, the low F1 scores during some epochs are intriguing. Are they caused by low precision, low recall, or both?

Response:

We thank the reviewer for this observation.

First, we replaced the term "vanilla CNN" with "baseline CNN" and now define it explicitly. The baseline CNN uses the same convolutional architecture as the SE-CNN but omits the squeeze-and-excitation blocks. It is used only as an architectural comparator; the SE-CNN is used for full-image inference.

Second, we examined the precision and recall curves separately to understand the temporary F1 decreases. The largest drops in validation F1 were accompanied primarily by decreases in validation recall, whereas precision changed less strongly. Because the validation set contained 194 patches, a relatively small number of threshold-crossing cases can produce visible fluctuations in recall and F1.

We have added this interpretation to the Results section.

Detailed comments: Results and Discussion—Formal

Reviewer 1:

Figures 5–8 use red and green, which are not ideal for readers with colour vision deficiencies.

Response:

Corrected. All revised figures use the Okabe–Ito palette: blue for SE-CNN, orange for manual, black for differences, bluish green for model-only, and vermillion for manual-only. PDF and 600-dpi PNG outputs are provided.

Reviewer 1:

Figures 4–6 are not explicitly referenced in the text.

Response:

Corrected. Revised Figures 4–6 are explicitly referenced in the Results section.

Detailed comments: Results and Discussion—Content

Reviewer 1:

The CNN outputs detection probabilities, and ripples above 70% probability are retained. Were the training labels probabilistic? If not, how should these probabilities be interpreted?

Response:

We thank the reviewer for this important observation. The training labels used in this study are binary ("ripple" and "non-ripple") rather than probabilistic. Consequently, the network output should not be interpreted as a calibrated probability. In the original manuscript we used the term "probability" too loosely. Throughout the revised manuscript we therefore replaced "ripple probability" with "model score" (or "SE-CNN score") and clarified that these values correspond to the output of the final classifier layer and are used only as a decision score for thresholding. Because no probability-calibration procedure (e.g., temperature scaling or isotonic regression) was performed, we do not interpret the output numerically as the probability that a ripple is present.

Reviewer 1:

Please compare the results with previous studies, including:

- Yue et al. (2010),
- Suzuki et al. (2011),
- Rourke et al. (2017),
- Wüst et al. (2026).

Response:

We thank the reviewer for this suggestion. We have expanded the Discussion to place the present results in the context of previous studies of ripple-like and small-scale wave structures.

Yue et al. (2010) is discussed in relation to seasonal and local-time variability of ripple occurrence in OH airglow observations from Colorado and Japan. Suzuki et al. (2011) and Rourke et al. (2017) are discussed as examples of studies at Antarctic sites that reported different temporal, directional, and seasonal characteristics. Wüst et al. (2026) is discussed in relation to machine-learning-based localization of spatially confined small-scale structures.

We emphasize that direct numerical comparison among these studies is limited because the observations differ in geographic location, emission layer, temporal coverage, instrument characteristics, preprocessing, structure definitions, score thresholds, occurrence denominators, and evaluation units. The revised Discussion therefore uses these studies to provide methodological and observational context rather than treating their numerical results as directly interchangeable with the present validation results.

Reviewer 1:

Figure 5

- Yue et al. (2010) and Wüst et al. (2026) do not report a pronounced autumn/winter maximum. How do these differences compare with the present results?
- The CNN detects more ripples than manual identification only during autumn. Can the authors explain this seasonal bias?
- Please define which months correspond to each season.

Response:

We thank the reviewer for these questions. We have clarified the seasonal definitions and revised the interpretation of the seasonal differences.

The revised manuscript defines spring as March–May, summer as June–August, autumn as September–November, and winter as December–February.

Within the selected manually positive validation cohort, autumn contained 380 automated tracks and 147 reference tracks, giving an automated-minus-reference difference of 233 tracks. The corresponding automated/reference count ratios were 2.59 in autumn, 1.47 in spring, 1.38 in summer, and 1.50 in winter.

The automated excess is therefore largest in autumn within this validation cohort, but it is not restricted to autumn. We do not interpret this as evidence of a physical seasonal mechanism because the analyzed frames were selected from manually positive observations rather than from complete observing nights, and the counts are not normalized by usable observing exposure.

The revised Discussion therefore considers possible methodological contributors, including differences in image background and quality, training-data composition, reference-annotation practice, and fragmentation or duplication in temporal linking, while explicitly avoiding attribution to tides, winds, gravity-wave filtering, instability, or secondary-wave generation without appropriate auxiliary observations.

Reviewer 1:

Figure 6

- This figure is not described in the manuscript.

Response:

We thank the reviewer for pointing this out. Figure 6 is now explicitly introduced and described in the Results section. The revised figure presents the seasonal distributions of provisionally linked reference and automated track durations. We emphasize that

these durations are products of the stated temporal-linking procedure and are not independently verified physical event lifetimes. The Results and Discussion also clarify that the distributions are conditional on the selected manually positive validation cohort and are not exposure-normalized seasonal occurrence statistics.

Reviewer 1:

Figure 7

- Manual detections do not exceed 5 minutes, making the statement that both distributions are similar somewhat misleading.
- Ripples are generally expected to persist for about 10 minutes. Why are some events much longer?
- The claim that manual detections show a higher fraction of long-duration winter events is not obvious from the figure.

Response:

We thank the reviewer for identifying this issue. We agree that the former comparison of manual and automated lifetime distributions was misleading.

The manual catalog does not contain independently defined event start and end times. The earlier plotting procedure assigned manual frame annotations to the first duration bin, so the absence of manual events longer than 5 min in that figure was an artifact of the plotting procedure rather than an observational result.

We have therefore removed the former interpretation of the manual and automated distributions as independently measured event lifetimes. In the revised analysis, both reference and automated detections are linked using the stated provisional temporal-linking procedure, and their resulting track-duration distributions are presented descriptively. We explicitly note that these tracks are algorithmically constructed and are not independently verified continuous physical events. Longer tracks may therefore reflect persistent structures, merging, or repeated association of nearby detections.

Accordingly, we no longer claim that the reference and automated duration distributions are generally similar or that the reference catalog contains a higher fraction of long-duration physical events in winter.

Reviewer 1:

The discussion attributes seasonal ripple occurrence to atmospheric background conditions, gravity-wave activity, and tides. However:

- Ripples may also be secondary gravity waves.
- Other studies report different seasonal behaviour.
- Evidence supporting claims regarding tidal and gravity-wave activity should be provided.

Response:

We agree with the reviewer. We have revised the Discussion to separate the observed seasonal differences in the image-based catalog from possible physical explanations.

The revised manuscript explicitly states that ripple-like structures may include secondary gravity waves as well as instability-related structures, and that the present airglow-only observations cannot distinguish these mechanisms. We also removed unsupported causal statements attributing the autumn difference uniquely to tides, background winds, gravity-wave filtering, Richardson number, convective instability, or secondary-wave generation.

The seasonal differences are now presented as properties of the selected validation cohort and detection procedure. We discuss possible methodological contributors, but emphasize that distinguishing among physical explanations would require complementary wind and temperature observations.

Reviewer 1:

The event recovery rate is stated to be 90%, but this metric is not defined.

Response:

We thank the reviewer for identifying this problem. We agree that the previously reported "90% event recovery" was not adequately defined and should not have been presented as a validated event-level performance metric.

We have therefore removed the 90% event-recovery claim from the revised manuscript. Instead, the present validation reports quantities whose definitions are explicit: 72.0% frame-level spatial recovery (299 of 415 evaluable manually positive frames) and 58.0% manual-object recall (324 of 559 eligible reference supports) based on maximum-cardinality one-to-one spatial matching.

These measures are defined in Sect. 2.2.4. We also explicitly state that archive-wide precision, event-level recovery, exposure-normalized occurrence rates, and independently verified reference–automated lifetime agreement are not established by the present validation design.

Reviewer 1:

I recommend changing the title of the final section from "Discussion" to "Discussion and Summary."

Response:

We thank the reviewer for this suggestion. The final section title has been changed from "Discussion" to "Discussion and Summary."

We again thank reviewer for comments that substantially improved the methodological clarity and scientific caution of the study.