Response to Reviewer 3 Comments

Thank you for your letter commenting on our manuscript entitled "***Innovative Cloud Quantification: Deep Learning Classification and Finite Sector Clustering for Ground-Based All Sky Imaging***" (MS No.: egusphere-2024-678). These comments are valuable and very helpful for the revision and improvement of our paper. We have carefully studied and made corrections, and hope to get your approval. The main changes of the paper and the responses to the review comments are as follows.

Comments 1: *In the title, abstract and introduction of the manuscript, "cloud quantification" has appeared many times, but without a clear definition.*

Response 1: Dear reviewers, Thank you for your valuable comments on our manuscript, in which you pointed out that we have mentioned the term "cloud quantification" several times in the title, abstract, and introduction without defining it clearly. We appreciate your careful review and constructive feedback. We will add a clear definition of "cloud quantification" in the introduction of the revised manuscript to ensure that readers can accurately understand the importance of the concept and its role in climate research. The changes are summarized below:

"Cloud quantification is the precise analysis of sky images to transform cloud body characteristics into a series of quantifiable parameters, including but not limited to cloud amount and cloud type, which are essential for understanding and modeling the Earth's radiation balance, energy transport, and climate change."

We hope that the revised manuscript will better introduce the concept of "cloud quantification" to our readers, and thank you again for your careful review.

Comments 2: *In the second paragraph of the introduction, the structure read somewhat chaotic. The classification method and observation instruments are also mentioned in the overview of cloud classification method. In the later description of the cloud quantification method, the advantages and disadvantages of the existing cloud quantification methods are not specifically introduced. It is suggested to reconsider the structure of this part.*

Response 2: Dear reviewers, we agree with your comments about the slightly confusing structure of the second paragraph of the introduction, the mixing of classification methods and observation instruments in the overview of cloud classification methods, and the failure to introduce the strengths and weaknesses of existing cloud quantification methods in the subsequent description of cloud quantification methods, and will conduct a comprehensive reorganization and optimization of the content of this part of the paper.

In the revised version, we will set up a separate subsection to systematically analyze the strengths and weaknesses of various cloud quantification methods, and adjust the overall structure of the introduction section to ensure that it starts from the climatic impacts of cloud phenomena to the scientific significance of cloud classification and quantification, and gradually transitions to the problems of the current technological tools and the methods proposed in this study to address these problems. It is hoped that the revision will enhance the coherence and hierarchy of the exposition, and allow readers to grasp the background and innovation of the study more clearly. The additions are as follows:

"The advantages of traditional image processing techniques are mainly reflected in the easy operation and low computational cost, which are suitable for rapid preliminary identification of cloud cover areas, however, the high sensitivity of such methods to changes in lighting conditions leads to unstable identification results under complex lighting dynamics, especially in the identification of high-altitude thin cirrus clouds, complex boundary cloud bodies, and multiple clouds, due to the lack of adaptive ability and accurate feature expression, it is difficult to achieve the ideal quantization accuracy and weak adaptability to atypical cloud types, which affects the accuracy of cloud calculation. Deep learning methods can efficiently and accurately classify and segment cloud images under complex cloud types and various lighting conditions by means of a deep neural network model driven by large-scale training data, and significantly improve the quantization performance under unfavorable lighting environments by combining with algorithms such as image enhancement. Deep learning methods also have obvious shortcomings, such as relying on a large amount of labeled data, high-performance computational resources, and the recognition performance in extreme lighting scenarios such as extremely bright or dark still needs to be improved. "

Comments 3: *In the abstract, the traditional NRBR recognition method may be just summarized in the introduction. In the last paragraph of the introduction, it mentioned the problem of cloud identification in current algorithm, but it is not clear which specific method the author referred to?*

Response 3: Thank you for your valuable comments, and based on your suggestion, we have, in the abstract section, in order to keep the content focused and compact, I have moved the specific description of the traditional NRBR recognition approach to the appropriate place in the introduction section so that the reader can have an overview understanding of this basic recognition approach before entering the main text. The limitations of the current algorithms for the cloud recognition problem, mentioned at the end of the introduction, have now been clearly referred to and detailed. The last paragraph of the original introduction has been modified as follows:

"Currently, many cloud recognition algorithms face significant challenges in dealing with different cloud types, especially high-altitude thin cirrus and transitional hybrid clouds (Ma et al., 2021). Among them, the traditional NRBR (Normalized Red/Blue Ratio) identification method, although able to provide preliminary cloud estimation in general, shows obvious

limitations in terms of shadowing effects and identification of thin cirrus edges due to the fact that it relies only on color features to make judgments, and the variation of illumination conditions greatly affects the identification results."

We will also further detail the recognition limitations of the existing methods in different scenarios in the subsequent chapters to ensure that the thesis is clearly and accurately presented. Thank you again for your help and support in improving the quality of the thesis. If you have any other suggestions or questions, please feel free to continue your guidance.

Comments 4: *In Section 2.3, how are the four types of data divided in image dataset?*

Response 4: Dear reviewers, thank you for your valuable comments on the paper. Regarding your question in Section 2.3 about how the four categories of data in the image dataset are categorized, I have made the following changes to the description in the original paper:

"We started by dividing the 4000 rain- and snow-free, unobstructed, high-quality all-sky images into four categories of 1000 images each, which are: cirrus, sunny, cumulus, and stratocumulus clouds; it should be emphasized that the division of clouds into the four main types is done here in order to accurately quantify the proportion of clouds in each category, rather than considering mixed clouds. These four cloud types play an important role in the region's weather and are the main references for this categorization."

If you have any other suggestions, please feel free to let us know and we will make more detailed notes or changes accordingly. Thank you again for your review and guidance!

Comments 5: *In Section 3.2.1, the description of neural network design is not clear. What are the advantages of the YOLOv8 architecture in solving the research problems of the current manuscript, and why choose the framework? In addition, YOLOv8 involves the convolution part, in which the process has certain requirements on the size of the input image. Why is the input image size of 680×680 selected in this paper?*

Response 5: Thank you for your interest in the neural network design section of subsection 3.2.1 of the article and for your valuable suggestions. We have enhanced the description of the advantages of the YOLOv8 architecture in solving the current research problem, and the following modifications and responses are made to address the issues you raised:

"The main reason why YOLOv8 is the preferred framework in this study is its unique design that can effectively handle the task of all-sky image cloud classification under complex lighting conditions. Compared to the previous YOLO series and some other classical image recognition models, YOLOv8 is able to extract richer gradient flow information by adopting Darknet-53 as the Backbone and utilizing the modified C2f module to replace the original C3 module (Li et al.,

2023), which is conducive to capturing the cloud's delicate textural and boundary features. Meanwhile, the PAN-FPN structure of YOLOv8 achieves model lightweighting while retaining the original high-performance performance, while the detection head part adopts a decoupled structure, which is responsible for the classification and regression tasks, respectively (Xiao et al., 2023), and adopts the binary cross-entropy loss (BCE Loss) for the optimization of the classification task, together with the distributed focus loss (DFL) and the complete IoU loss (CIoU) for bounding box regression prediction, this detection structure can significantly improve the detection accuracy and convergence speed of the model (Wang et al., 2023)."

Regarding the selection of the input image size, we set it to 680 × 680 pixels, which is because the convolutional part of the YOLOv8 network does have some requirements on the input image size. This size was selected based on the consideration of several factors:

(1) The original all-sky image has extraneous black background as well as feature interference, and the resolution of 680 × 680 is the result of compressing the image after removing the image edges in a specified range of zenith angles.

(2) The resolution of 680 × 680 retains the main detailed features of the clouds in the image while significantly reducing the file size, which is beneficial to the loading and computational efficiency of the model in the training phase;

(3) This size not only meets the demand of YOLOv8 network structure on the image input size, but also takes into account the various morphological features of the cloud body in the image, which ensures that the model is able to maintain good recognition performance when dealing with cloud bodies at different scales and under complex lighting environments.

Thank you again for your review and questions, and we hope that these improvements will shed more light on the applicability of the YOLOv8 architecture in this study and the reasonableness of the selected input image sizes. Please feel free to give us feedback if there are any other areas that need further clarification or refinement.

Comments 6: *For what reason is the training epoch set to 400? Because in Figure 4, when the epoch is greater than 200, it is found that F1 is basically unchanged, and the loss is no longer reduced.*

Response 6: Thank you for your valuable comments on the experimental parameter setting section. Regarding the reason for setting the number of training rounds to 400, we did observe in the initial experimental phase that the training process stabilized the F1 scores after about 200 epochs, and the reduction of the loss function decreased. However, setting 400 epochs is mainly based on the following considerations:

(1) Global optimality exploration: although the model performance metrics are no longer significantly improved after 200 epochs, we note that the longer training period helps the model to jump out of the local optimal solution and search for possible better solutions, which may be beneficial to the model's generalization ability and stability even if the gain is small.

(2) Avoiding the risk of early stopping: stopping training early may lead to fluctuations in the model's performance on the validation set, and the choice of 400 epochs is intended to ensure that the model learns the diversity and complexity of the dataset adequately over a sufficiently long period of time, and to prevent potential performance improvement opportunities from being missed by ending the training too early.

(3) We considered the possible risk of overfitting, as well as the improvement in accuracy and precision. Ultimately, 400 epochs were chosen as the default training termination point, and doing so yielded better classification results in this study.

Combining the above reasons, we decided to set the number of training rounds to 400, which contributes to the robustness of the model even though the growth of the performance metrics slows down during the later stages of training. Meanwhile, we also noticed your question about the performance bottleneck at the late stage of training, and in our future work, we will consider introducing more refined training strategies, such as Early Stopping or other optimization methods, to save computational resources and avoid overfitting. Thank you again for your review and suggestions, and we will reasonably adjust and optimize the experimental scheme according to the actual situation.

Comments 7: *In the description of the evaluation metrics in Section 3.2.3, what are the validation set and test set mentioned? What the descriptions such as true positive and false negative represent? Please give more clear explanations.*

Response 7: We often thank you for your valuable suggestions on the evaluation metrics part of our study. In the revised Section 3.2.3, we have elaborated the concepts of validation and test sets and their roles in evaluating model performance:

In the field of machine learning and deep learning, datasets are usually categorized into three parts: training set, validation set, and test set. The training set is the part used to train the model to learn the intrinsic laws of the data; the validation set is used to adjust the model parameters during the training process and test the model generalization ability to determine the best model; and the test set is completely independent of the training process, and is only used to evaluate the final performance of the model on unknown data after the model training is completed to ensure that the evaluation results are fair and objective.

For the description of True Positive (TP) and False Negative (FN), we have explained these concepts and how they are calculated in detail in the " 3.2.3. Cloud Classification Evaluation Indicators " section of the revised draft. We have explained these concepts and their calculations in detail in the "3.2.3:

"True Positive (TP) denotes the actual number of positive samples that the model correctly predicts as positive category (i.e. cloud category), which represents the number of real cloud images that the model successfully recognizes. False Positive (FP) denotes the number of samples that the model incorrectly predicts as positive category but actually belongs to the negative category (non-cloud category), which implies the number of cloud images that the model misidentifies. False

Negative (FN) denotes the number of samples that the model incorrectly predicted as a negative category but actually belonged to a positive category, which represents the number of cloud images that the model failed to identify. "

By clarifying these concepts, we have ensured a clearer and more thorough presentation of the evaluation metrics section to facilitate the reader's understanding of how we quantitatively assessed the performance of the cloud classification model on the validation and test sets.

Comments 8: *In part 3.3, how to estimate the A value and how to reflect the adaptive process in the defogging algorithm? Do you mean that each type of cloud selects a different A value to achieve adaptive?*

Response 8: Thank you very much for your interest and guidance on the atmospheric light estimation and its application to adaptive defogging algorithms in Section 3.3 of the paper. In response to your questions, I will make a more explicit and detailed explanation in the revised text.

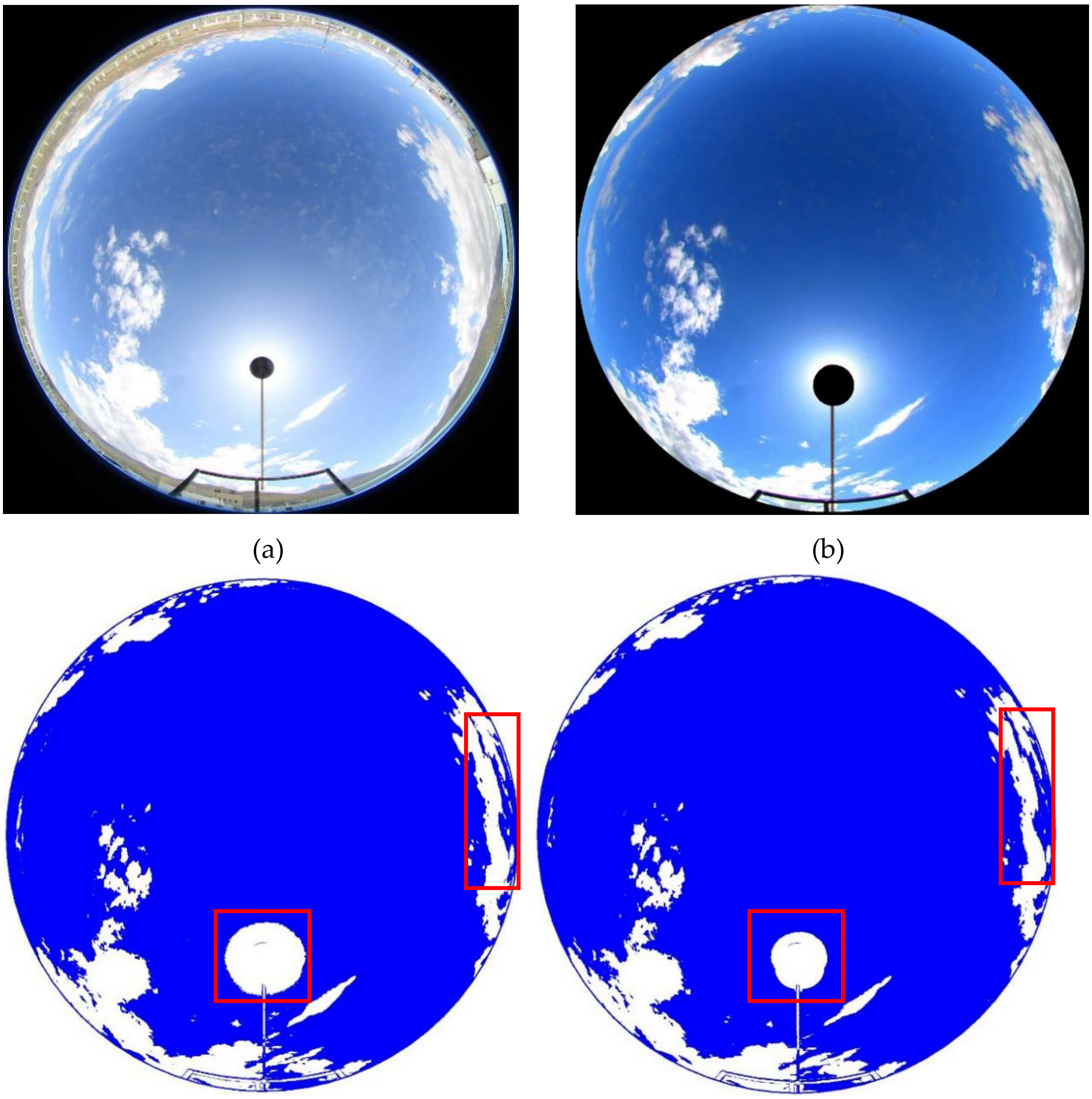
In our defogging algorithm, the A-value represents the global atmospheric light intensity, which plays a key role in the defogging effect of the whole image. In the dark-channel a priori algorithm proposed by Kaiming et al. (2009), we first calculate the minimum values of the three RGB color channels at each pixel point to construct a dark-channel image. Then, we use the non-zero minimum value in the dark-channel image to estimate the global atmospheric light intensity A. This value reflects the atmospheric light intensity in the scene that is not shaded by clouds and is directly illuminated by sunlight.

"In the image enhancement algorithm, the atmospheric light value A directly impacts the intensity of dehazing. Thanks to the powerful cloud classification network, we design adaptive enhancement strategies after recognizing different cloud types. For relatively thin cirrus, excessive enhancement may filter them out, hence smaller A values are chosen to preserve details. "

We have designed a strategy to dynamically adjust the A-value according to the morphological characteristics of the cloud body and the complexity of the lighting conditions. For different types of clouds, we will flexibly choose the most suitable A-value to optimize the de-fogging enhancement effect, which in turn improves the recognition accuracy of cloud edges and thinning regions. Thanks again for your review, we have added the above content in the revised manuscript to fully elaborate the estimation method of A-value and its dynamic adjustment process in the adaptive defogging algorithm.

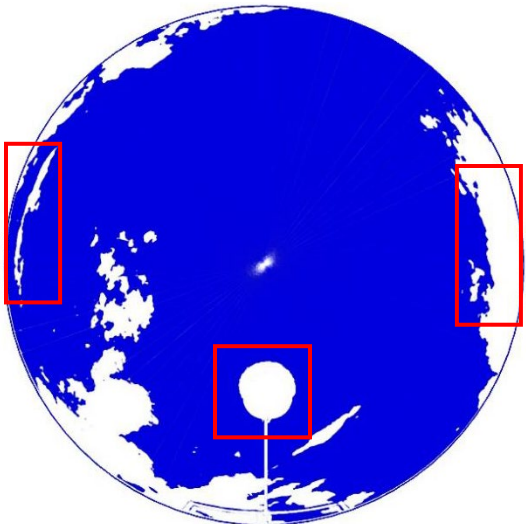
Comments 9: *In figure 7, it can be seen that the contrast between cloud and clear sky is obviously enhanced after image enhancement. I am curious about whether the NRBR method is applied to the images before and after enhancement, and will the results be different? Please provide the cloud detection results of the NRBR method for the enhanced image and compare it with your new method.*

Response 9: Thank you for your interest in the comparative effect of image enhancement before and after and the effectiveness of NRBR method applied on enhanced images. On the basis of our original research, we will compare the performance difference of NRBR method before and after image enhancement, firstly, we have processed the original image with adaptive enhancement, and then we apply the NRBR method and our new method of finite sector segmentation combining with K-means clustering to cloud detection of enhanced image respectively, and we get the results shown in Figure 1:



(a) (b)

(c) (d)



(e)

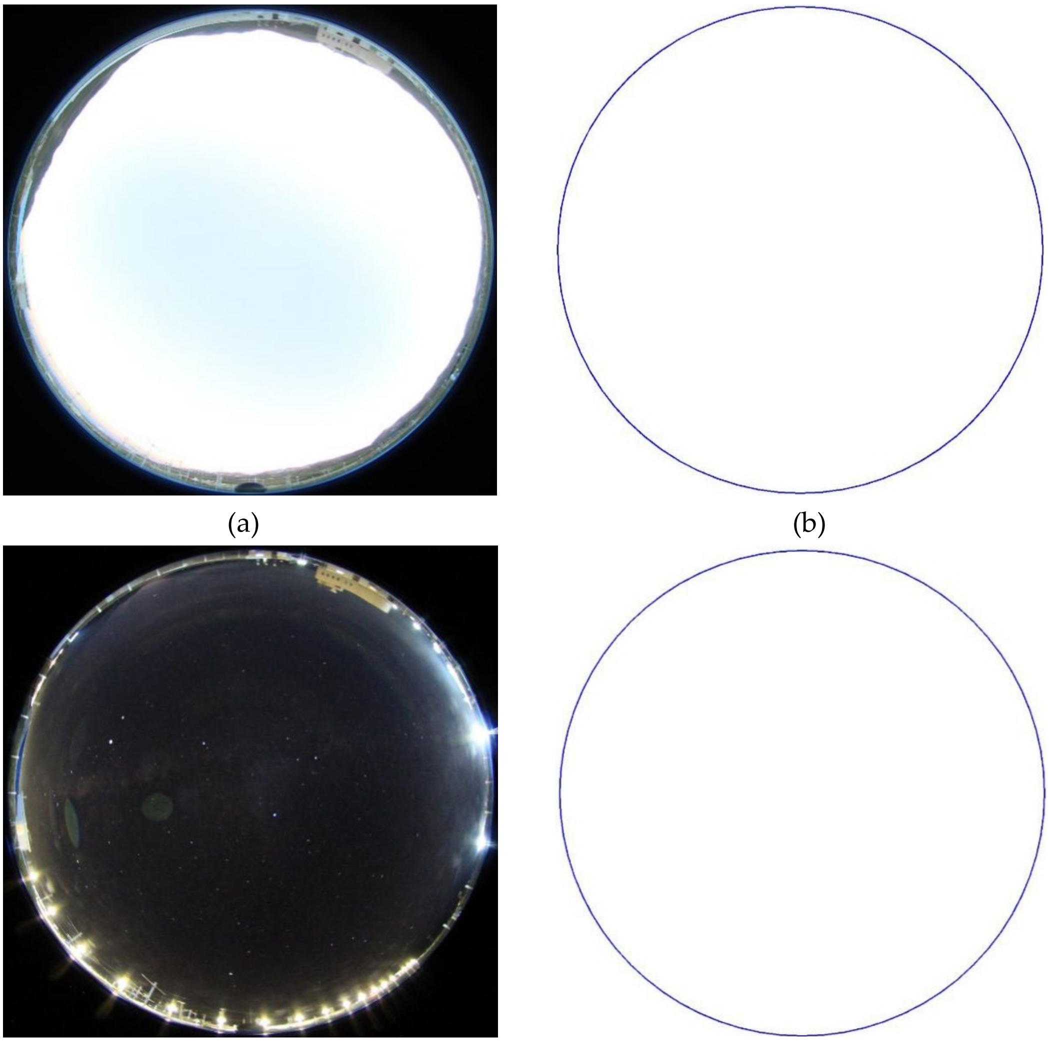
Figure 1. Cloud recognition effect comparison image. (a) Original cloud image; (b) The result after image enhancement; (c) Recognition effect of non enhanced images using NRBR method; (d) The recognition effect of the enhanced image using the NRBR method; (e) Finite sector segmentation k-means clustering results.

As shown in Figure Figure 1d, on the enhanced image, the NRBR method is somewhat improved in recognizing complex lighting conditions, especially the overexposed region (around the sun) and the cloud bottom details (on the right side of the image), but it is still not as good as our proposed novel method. The new method is able to recognize cloud boundaries (e.g., thin cloud on the left side and thicker cloud bottom on the right side of Figure 1e) more accurately and reduce confusion and misclassification when processing the enhanced image, especially in the sunlight-adjacent region with strong illumination, the edge of the thin cloud, and the bottom of the thicker cloud, where the new method shows a higher recognition accuracy.

Comments 10: *The algorithm considers using images from 9 : 00 to 16 : 00 in the day as training data, and the paper mentioned that the illumination has a great interference to the recognition results. What are the identification results before 9 : 00 and after 16 : 00 ? Are there any examples? In the cloud cover time series of Fig.8, you give the comparison between the proposed algorithm and the traditional algorithm. How do you calculate the improvement in accuracy of the new method compared to the traditional algorithm?*

Response 10: Thank you for your insights and specific questions about my research. In response to your questions, we add the following:

Regarding the results of image recognition for time periods other than 9:00 to 16:00 during the daytime, we did notice that image recognition outside of these two time points becomes more difficult and prone to misclassification due to the significant effect of lighting conditions on cloud recognition. In our initial experiments, we found that the overall whiteness of the sky could not distinguish between blue sky and white clouds in the morning and evening hours due to the low sun angle, complex illumination and large intensity variations, leading to a decrease in the accuracy of cloud recognition. Moreover, due to the limitation of the sensor, the blue sky and clouds at night cannot be directly photographed by visible light, and need to be detected with the help of other data. We have shown the individual images before 9:00 and after 16:00 in Figure 2 below:



(a) (b) (c) (d)

Figure 2. Cloud detection results before 9 a.m. and after 16 p.m. (a) Example of an all-sky image in the morning, just before 9 a.m., when the sun comes ups; (b) The cloud identification result of Fig. 2a; (c) Individual examples of all-sky images after 16 p.m.; (d) The cloud identification result of Fig. 2c.

Regarding the cloud coverage time series comparison shown in Fig. 8 of the original manuscript, we used the cloud amount calculation results of the two algorithms (i.e., the algorithm proposed in this study and the traditional NRBR algorithm) on the same set of image data and quantitatively analyzed them by comparing the accuracy of the cloud amount identification at 15:00 PM every day. We calculate the degree of match between the cloud amount identified by the new method and the actual naked-eye observation records in the image-enhanced data and compare it with the results obtained by the traditional NRBR method without enhancement. The improvement in cloud identification accuracy of the new method is demonstrated by the statistically significant percentage improvement in accuracy and the reduction in the false positive rate under specific lighting adverse conditions such as the bottom of the cloud cover and overexposed regions around the sun.

Comments 11: *The classification accuracy of the algorithm is very high. But I 'm interested in misclassified images. Can we know the specific reasons for these misclassifications?*

Response 11: Thank you very much for recognizing the classification accuracy of the algorithms in this paper, as well as expressing interest in the potential causes of misclassification. In the course of our research, we found that the identification of hybrid clouds is one of the main causes of misclassification. Hybrid clouds make it challenging for the algorithm to process such images due to their complex internal structure, the difficulty in defining the transition region of different cloud types, and the variation of cloud types under lighting conditions. In addition, complex lighting dynamics, especially in intense lighting as well as backlit scenes, may exacerbate the blurring of cloud body boundaries and features, further affecting the classification performance. We have examined the misclassified images output from the model one by one and found the following common problems:

(1) Complexity of hybrid cloud layers: hybrid clouds have diverse morphologies and complex internal structures, with both cumulus and laminar cloud elements, and this combination of features sometimes makes the model hesitant to make decision boundaries.

(2) Influence of lighting conditions: Under too strong or too weak lighting, the visual characteristics of cloud layers may be significantly distorted, leading to model difficulties in identification.

(3) Blurred edges of cloud bodies: the boundaries between cloud bodies and the sky are not clear in some images, which makes the model prone to confusion when performing segmentation and classification.

(4) Transitional clouds: Transitional patterns produced during the formation and dissipation of clouds are often difficult to categorize with typical cloud type features.

In future versions, we will strengthen our research on the cloud recognition problem under hybrid clouds and complex lighting conditions, and try to improve the accuracy of hybrid cloud recognition by improving the model architecture, optimizing the image preprocessing techniques, and introducing more hybrid cloud samples to train the model.

Comments 12: *In the third paragraph of the introduction, the sentence "Some studies train k-means models to swiftly cluster and recognize cloud and clear sky regions in whole sky images, improving cloud quantification speed and efficiency." lacks literature citation.*

Response 12: Thank you for your valuable comments on the paper.In the third paragraph of the introduction section, regarding the description related to the use of k-means model for fast clustering and identification of clouds and clear sky regions in all-sky images to improve the speed and efficiency of cloud computation, you have pointed out that there is a lack of literature citation.We apologize for this and will fill in the academic basis for this section in the revised manuscript.

"Krauz and other research teams have previously successfully analyzed all-sky images using the k-means clustering algorithm to quickly and efficiently delineate cloud cover and clear-sky regions, significantly improving the speed and efficiency of cloud quantification tasks (Krauz et al., 2020)."

Thank you again for your professional guidance, which certainly helps us to improve the quality of our dissertation.

Comments 13: *When the "Yangbajing Comprehensive Atmospheric Observatory" appears for the first time, it is recommended to give specific latitude and longitude coordinates.*

Response 13: Thank you very much for your valuable suggestions. We have provided specific latitude and longitude coordinates for the first reference to the "Yangbajing Comprehensive Atmospheric Observatory" in the subsequent manuscript to enhance the reader's clarity in locating the site, with the following modifications:

"The Yangbajing Total Atmosphere Observatory (90°33′E,30°05′N) is located next to the Qinghai-Tibet Highway and Qinghai-Tibet Railway, 90 kilometers northwest of Lhasa, Tibet, in an area with an average elevation of 4,300 meters."

Comments 14: *In Section 2.3, pay attention to the number of samples, it should be 15 samples per day instead of 16.*

Response 14: Thank you very much for your careful review and correction of the sample size of data in this paper.In section 2.3, as you pointed out, we did incorrectly describe the study as collecting 16 samples per day. In response to your comment, we have checked the raw data and confirmed that 15 samples were actually collected each day. We have amended the relevant paragraph to read:

"Considering that images during sunrise and sunset hours are susceptible to lighting conditions, we only select images between 9am and 16pm hours each day. Also, to reduce the correlation, only one image is selected every half hour, which results in 15 sample images per day."

Comments 15: *In Part 2.2, what is CMOS? If it is an abbreviation, please give a full name.*

Response 15: Thank you very much for your valuable comments on my article. I apologize for the inconvenience caused to the readers by using the acronym CMOS instead of giving its full name in part 2.2 of the article.CMOS stands for Complementary Metal-Oxide-

Semiconductor, a technology widely used in the fabrication of digital image sensors, especially for capturing high-resolution sky images in the all-sky imaging systems covered in this paper. In the revised version, I will give a full explanation of the terminology when CMOS is first mentioned to ensure that the content is clear and understandable.The revised content is as follows:

"This visual imaging device is equipped with a complementary metal oxide semiconductor (CMOS) image sensor system with an ultra wide angle fisheye lens design, which can regularly capture visible light spectrum images across the entire sky range; The integrated sun tracking system can accurately calculate and track the position of the sun in real-time, ensuring effective blocking of direct sunlight shining into the CMOS system, thereby protecting its sensitive photosensitive components from damage and significantly reducing the interference effect of white light around the sun on subsequent image processing."

Comments 16: *Figure 8 in "Likewise, the cloud cover recognition effect at the base of heavier clouds and the overexposed area surrounding the sun are greatly improved when the image is enhanced, as seen in Figure 8g,h." should be changed to Figure 7.*

Response 16: Thank you for meticulously reviewing and pointing out a figure citation error in the paper. You pointed out that when describing the improved cloud recognition effect in section 4.2 of the paper, the figure number cited should be Fig. 7 instead of Fig. 8. I apologize for this and have immediately corrected the manuscript. The revised presentation is as follows:

"Similarly, after image enhancement processing, as shown in Figures 7d and 7h, the cloud recognition effect at the bottom of thicker cloud layers and overexposed areas around the sun was significantly improved compared to Figures 7c and 7g."

In subsequent revisions, we will be more careful in verifying all references to charts and tables to ensure that the content is presented correctly.

Comments 17: *Some grammatical words in the article need to be checked carefully.*

Response 17: We take your comments about the grammar of the article very seriously, and we apologize for the poor grammar in the manuscript. We have spent a long time revising the manuscript, studying the language and readability, and involving professionals in correcting the language, correcting and optimizing the parts of the text that may contain grammatical errors, and trying to eliminate any barriers to reader comprehension.

We ask for your additional guidance and suggestions as you review the new revised manuscript. We firmly believe that after this comprehensive proofreading and revision, the linguistic quality of the article will be enhanced, thus better serving the dissemination and communication of scientific research. Thank you again for your attention and support to this article!