Dear Reviewer,

We are grateful to your constructive comments and valuable suggestions, which helped us to further improve our manuscript and dataset. Below we address your concerns point-by-point, with the original comments in **black** and our response in **blue**. The revised sentences in the manuscript are indicated in ***italics***.

Overview:

Marine low-clouds cover the majority of the ocean, and play an important role on the Earth's radiation budget. Due to a lack of local or ground-based observations, satellites become powerful tools for MLC measurement, while satellite observations over nighttime are still relatively limited. Thus, this study by Wu et al. introduced a deep-learning based method for the classification of MLC and their mesoscale morphology using MODIS observations, and a global dataset is developed as well. Both all-day model and day-time model were developed and evaluated. It is interesting to find some differences on the daytime and nighttime MLC, and distinct seasonal variations were also noticed for different MLCs. The new method as well as the resulting dataset is an important addition for the community, and the paper is well organized and presented. The paper could be considered for publication after considering following suggestions.

Thank you very much for acknowledging our work and for your valuable comments and suggestions. We fully agree your concern on the construction of our dataset, the representativeness of training data, as well as the model training and validation process. In the revised version, we have further clarified the construction of our training, validation, and testing datasets, evaluated the data representativeness, and explained the reliability of our nighttime results. Please refer to the response below for further details.

1. The quality of the training and testing dataset has been essential for DL-based models, so the datasets for the training should be carefully constructed. The 2.2 Data session gave some information on the dataset, while missed some as well. For example, Figure 1 gave some examples of MLCs of different kinds, and how was the original training dataset classified? The independency of training and testing dataset is also important, so I would suggest to introduce the testing and evaluation dataset at the Data session as well.

We apologize for any confusion caused by the lack of detailed information. Our training, validation, and test datasets are all sourced from the same manually annotated dataset. While they originate from the same data pool, they have been randomly partitioned into mutually independent subsets to ensure robust evaluation and model generalization. To clarify further, we have made some modifications to the original text, and it currently reads like: *"A total of 38,756 labeled daytime scenes were obtained, including 3,548 scenes of solid stratus, 6,277 of closed MCC, 3,345 of* *open MCC, 6,739 of disorganized MCC, 8,947 of clustered Cu and 9,900 of suppressed Cu. These scenes were then randomly partitioned into three mutually independent datasets for training, validation, and testing, with a distribution ratio of 3:1:1 respectively."* (Lines 129-132)

2. Cloudy and atmospheric properties show clear seasonal variations. For example, surface and atmospheric temperatures may significantly different from season to season, and this is also true for clouds. It is mentioned that only the results over the first half of 2014 were used for data training. Would such choice of results from half a year influence the DL performance?

Your suggestions are highly valuable. Indeed, the formation and development of clouds are significantly influenced by meteorological conditions, such as sea surface temperature and lower tropospheric stability (LTS), which differ from season to season. Nevertheless, as the ultimate manifestation of meteorological conditions, cloud patterns exhibit certain similarities across different regions and seasons. That is, the cloud patterns in a specific region resemble those on a global scale, and the cloud patterns in the first half of the year are similar to those throughout the entire year. Therefore, the cloud patterns contained in our dataset can largely represent the all-year and global clouds.

To validate our hypothesis, we examined the differences in the probability density functions (PDFs) of thermal radiance, cloud optical thickness (COT), and cloud morphology between our training dataset and a global full-year dataset. As shown in the Figure R1. The results revealed a substantial overlap between the two PDFs, suggesting that the training data we selected is relatively representative and can be used to substitute the global full-year dataset. Therefore, we have added a statement to the article: "*The representativeness of this dataset was validated as the probability density functions (PDFs) of thermal radiance data and cloud optical thickness show large overlap with those of the global and full-year dataset (Fig. S1)."* (Lines 120-122)

Despite this, in the future, we plan to re-label the global dataset for all seasons, both day and night, and update our model and products in subsequent iterations.

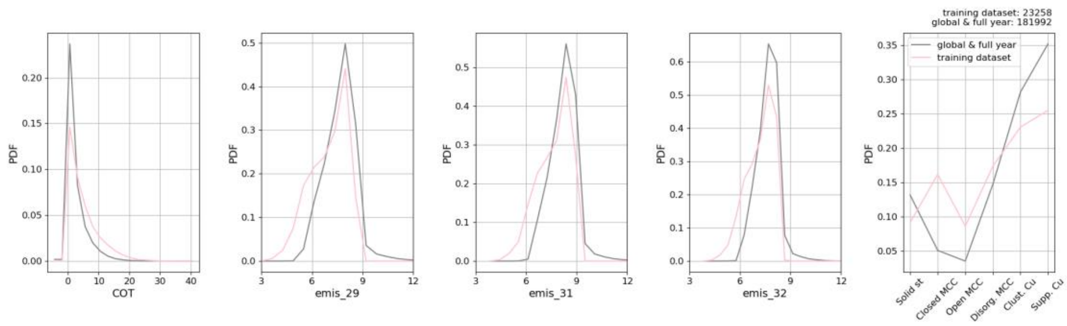

Figure R1. The comparison of probability density functions (PDFs) between our training dataset and a global full-year dataset. (a) PDFs of cloud optical thickness

(COT); PDFs of radiance data from infrared channels: (b) 29, (c) 31, (d) 32; (e) PDFs of cloud morphology.

3. Would it be possible to include the exact variables of input for different models in the flowchart of figure 2? This would be very helpful to better understood the details of the model efficiently.

Thank you for your valuable suggestion. We have included the exact input variables in Figure 2a, as also shown in the following Figure R2, to enhance the reader's understanding of the model details.

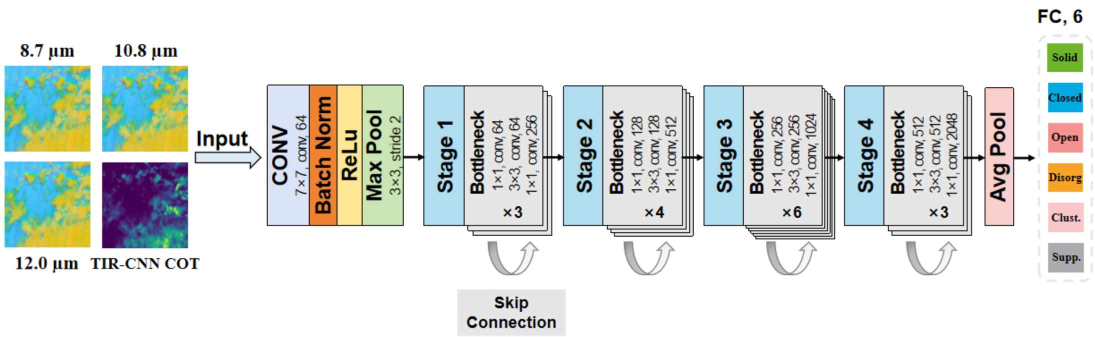

Figure R2. The revised ResNet-50 model structure figure.

4. The example of solid stratus show relatively regular linear structure, and are such structures natural? Please double check.

These linear structures are strip noise caused by the components of satellite sensor. Although present in the classification processes, they can be identified and filtered out by the model, thus having minimal impact on the training and classification results. We have attempted conventional methods, including mean filtering, Fourier transform, and directional filtering, to eliminate these strips and enhance the visual quality; however, none of them have proven effective yet (Figure R3). AI-based removal techniques appear promising, but additional time is required to fully master it.

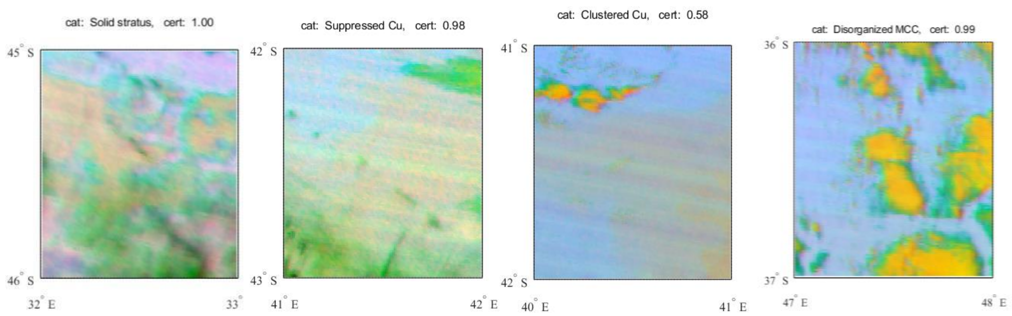

Figure R3. The images processed with directional filtering.

5. The training model based on daytime results is extend to nighttime observations. This is essential for the work, and could be tricky. The validation of the model for nighttime observations is very important, while only some examples were shown in Figure 4. Would it be possible to improve the validation to ensure the reliability of the results for nighttime?

We agree with your concern. The difficulty in nighttime predictions arises from the lack of cloud thickness information. However, as we responded to the Comments #1 from Reviewer #1, the COT retrieval method by Wang et al. (2022) has been proven reliable and can effectively replace MODIS COT for accurate nighttime classification (Figure R4). Moreover, since cloud classification requires pattern recognition rather than quantitative values, the differences in infrared radiance and COT value between day and night do not significantly affect the classification.

In addition, we examined the probability density functions (PDFs) of COT and thermal radiance between our training dataset (daytime) and a nighttime dataset (Figure R5), which present large overlap. The significant similarity between daytime and nighttime input data indicates less extrapolation by the model and ensures the reliability of our nighttime results. Therefore, we added some statements in the manuscript: *"In addition, we further examined the differences in the PDFs of the thermal radiance data and the TIR-CNN-based COT between our training dataset (daytime) and nighttime dataset. As depicted in Fig. S4, these PDFs nearly overlapped, which means less extrapolation will be introduced when the model is generalized to nighttime data. And it also illustrates the credibility of our nighttime classification results."* (Lines 145-148)

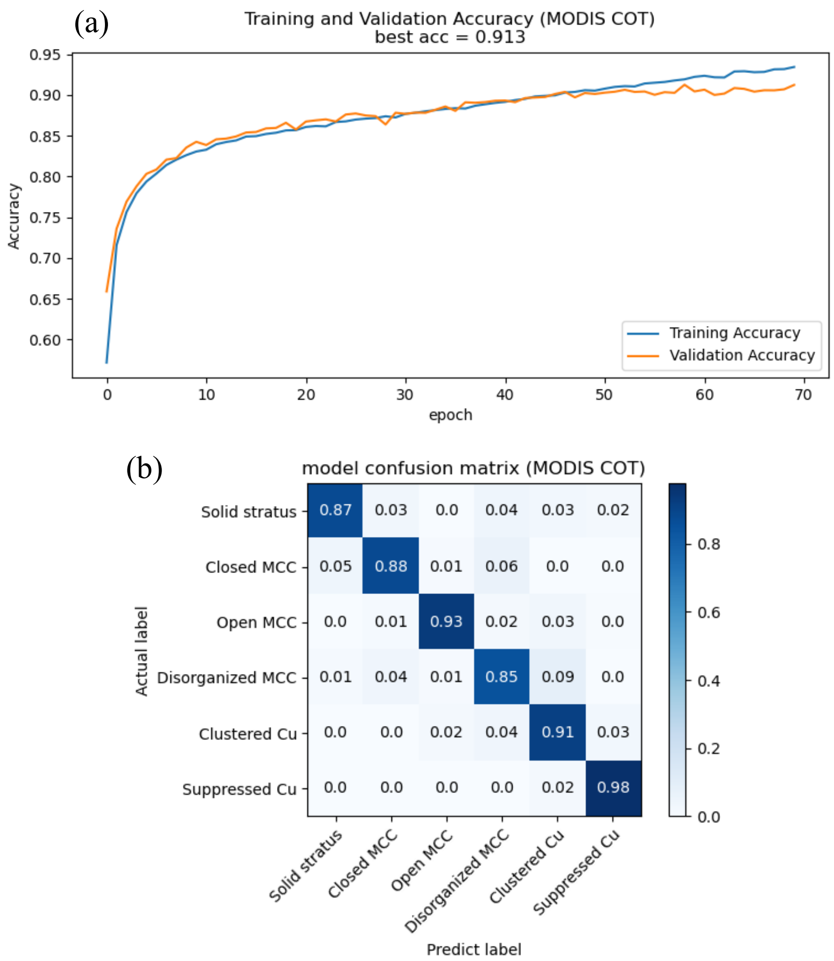

Figure R4. Model training results based on MODIS COT. (a) the model's accuracy on the training and validation datasets. (b) the confusion matrix of the model.

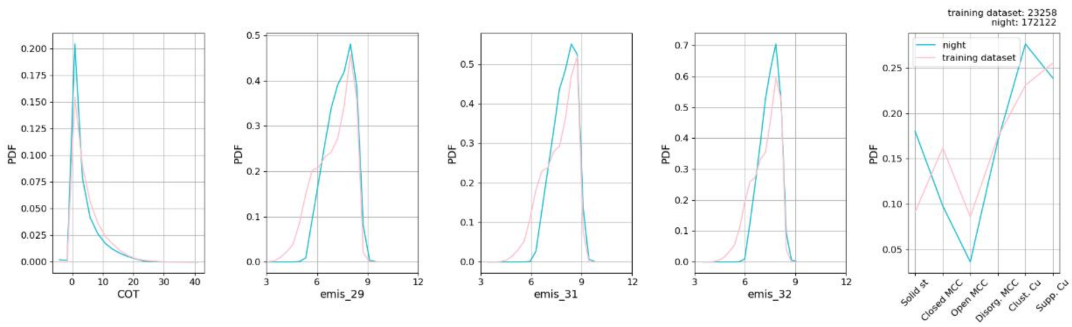

Figure R5. The comparison of probability density functions (PDFs) between our training dataset and a nighttime dataset. (a) PDFs of cloud optical thickness (COT); PDFs of radiance data from infrared channels: (b) 29, (c) 31, (d) 32; (e) PDFs of cloud morphology.

6. Figures 7 indicates clear day and time differences between RFO of different MLCs. Could the authors give some discussions on the reasons for the differences?

Thank you for pointing us in this direction. Your suggestion actually aligns with our own thoughts, and it is part of our next steps.

In this study, we have conducted a statistical analysis of six meteorological conditions in the article but found that the variations in these meteorological factors between day and night were not significant. Therefore, other factors, such as cloud-top radiative cooling, might be responsible for the observed differences. We plan to further investigate the controlling factors behind the day-night changes of cloud morphology in future work.

Furthermore, given that the primary focus of this article is to introduce the cloud dataset and the machine learning method, we feel that including an analysis of cloud-controlling factors might shift the focus away from the main theme.

References

Wang, Q., Zhou, C., Zhuge, X., Liu, C., Weng, F., and Wang, M.: Retrieval of cloud properties from thermal infrared radiometry using convolutional neural network, Remote Sensing of Environment, 278, 113079, https://doi.org/10.1016/j.rse.2022.113079, 2022.