Review of “Semi-supervised machine-learning method for analyzing images from the Balloon-borne Ice Cloud particle Imager B-ICI”

This study proposed a machine learning (ML) algorithm for ice crystal image segmentation and classification. It is claimed to be a semi-supervised algorithm by the authors. In-situ imaging techniques can measure microphysical properties of individual cloud particles, which are valuable to study the climate and local influence of cloud. The traditional image analysis relys on manual processing which are time-consuming and labor-intensive. While over past decades, various machine learning (ML) methods were being applied to detect and classify of cloud particles collected by different devices. However, there is not a universal model that can be applied across different instruments. The nature of the data made the authors propose a 2-step model for processing and classifying them. A segmentation model that is responsible for filtering ice particles from the background, and a classification model that is used to classify filtered ice particles to 4 prescribed categories. An achievement of the ML algorithm is that it was successfully applied on the data collected by B-ICI for the first time. Based on the evaluation on 200 particles. two major results were concluded in this manuscript.

1. The filtered ice particles by the segmentation model showed high correlation with the hand-labled ice particles in terms of the maximum dimension and particle area. The maximum size distribution of model output and ground truth showed good agreement for particle larger than 100 μm.
2. The agreement of columns, compacts and rosettes with ground truth was around 80%

Comments and questions

In general, while the research topic and proposed ML model are of significant interest, the major concerns regarding the presentation of this manuscript, the novelty of its contributions and the sufficiency of analysis and interpretations currently prevent me from recommending publication in its present form. I believe the manuscript has potential, but it would require substantial revisions to address several issues which will be detailed discussed in the following review script.

Incomplete literature review and vague statement of motivation/research questions

In the introduction section, the review of machine learning (ML) methods applied to ice crystal images is not complete and up-to-date. **Lindqvist et al. (2012)** and **Praz et al. (2018)** used Principal Component Analysis (PCA) and logistic regression for classifying ice crystal images before **Xiao et al. (2019)**. More recently, the field has seen significant advancements. **Schmitt et al. (2024)** proposed using a **Visual Geometry Group (VGG)** model for ice crystal classification, while **Zhang et al. (2024)** developed a specialized rotated object detection algorithm that can identify ice crystal aggregates down to the monomer scale. Additionally, **Chu et al. (2025)** introduced a semi-supervised algorithm to efficiently classify a large number of unlabeled images. This additional revision provides a more complete and chronological overview of the key studies, showing how the field has evolved from traditional methods to more advanced deep learning techniques.

The current statement about a "universal model" (lines 37-40) is a bit confusing and doesn't clearly align with the rest of the manuscript's apparent purpose. To improve this, the authors need to re-consider their research questions or motivation in the introduction.

Insufficient and imprecise description of data and CNN model

In Section 2.2, the description of image data is too brief. Although the authors cited the data analysis paper (Wolf et al. (2018, 2019)), in order to make it more "reader-friendly", the authors should provide a concise summary of the dataset directly in this section. Additionally, the authors should include a description of all datasets used in your study, including the one from the 2024 campaign, which was used for evaluation in Section 4. The summary of the dataset should include at least the name of the campaign and the number of particles. It would be better if the author can also describe the environmental conditions like in-cloud temperature and supersaturation briefly.

The model description (Section 3) lacks references. Most of the descriptions of CNNs are based on past work in the field of computer vision, while not much relevant work is cited here. In addition, the expressions are not accurate. For example, the expression: "The CNN processes this matrix by evaluating each pixel along with its neighboring pixels, assigning values to local pixel groups," is vague. "Assigning values" does not clearly describe how convolution between a kernel and the corresponding receptive field is calculated. I suggest the authors improve the description of the CNN to make it more precise.

Line 148, please cite the paper for those models: ResNet, VGGNet, and AlexNet.

Insufficient description of segmentation model training/evaluation

The current description of the segmentation model is too brief. The authors need to provide more details about its architecture, justify the choice of EfficientNet-B7, and clarify the training and evaluation process.

Most segmentation models follow an encoder-decoder structure. EfficientNet is just a backbone which can be applied to different tasks. It would be better if you revealed more information about the structure of the segmentation model and justified your choice of EfficientNet-B7 instead of simply mentioning "through repetitive training."

The description of the beta factor in Section 3.2 is indeed vague and incomplete. The author needs to explain why it's necessary for their dataset and provide more details about how it was used in their model training. For example, since the beta factor is used for balancing the FN and FP values, the reason why the authors want to balance them in their datasets should be specified here.

I am confused in lines 178-179 about why the performance on fresh data is inspected visually instead of evaluated according to some mathematical metrics. Why is human vision here more trustworthy than statistics? As far as I understand, the choice of the hyperparameter beta in this study is based on human vision by iteratively inspecting the trained segmentation model output of the validation set. Then the question is why not using a grid search method on the hyperparameter beta based on mathematical metrics?

Until Section 3.2, the process of training the segmentation model seems like a fully-supervised method to me. The training of the segmentation model starts from inputting an annotated dataset according to **F**igure 6. The further training of the model with unseen data is still based on human supervision. Therefore, the segmentation model of this manuscript cannot be defined as a 'semi-supervised' model.

Typo: line 160 "amountof"

Insufficient description of classification model training/evaluation

Line 204, it would be better if the authors could provide the number of particles used in training.

Line 213, typo "number"

Lines 214-219, the details of each component in the confusion matrix are not described. In Line 219, the decimals in Figure 9 are defined as 'confidence', which is vague to the audience what they represent. Normally, the decimals on the diagonal are the precision of each category. Plus, you mention the lowest one is 85%, which I did not find in Figure 9.

In general, the process of training the classification model is also fully supervised.

Insufficient model performance evaluation

The major problem here is that the segmentation model and classification model were evaluated only on a 200-particle dataset, which is too small to achieve a statistically significant conclusion, although the way of evaluation (size distribution comparison and confusion matrix) is good.

Lack of novelty

There have been many classification models developed for ice crystal images. Therefore, a simple 4-category classification model in this manuscript is not a novel contribution. However, the segmentation model developed for an oil-coated ice crystal image system could be a potential novel contribution to the community. The segmentation model has the potential to filter out ice particles from films among other artifacts. As mentioned before, the authors still need to add more analysis about the evaluation of the segmentation model.

Another problem is that the authors claimed the method is semi-supervised while the flowchart indicates that each node of training still involves human supervision.

Code and data availablity

I strongly suggest to make code and data available to public through associated GitHub/Zenodo repository even at the preprint stage.

Quality of figures

All figures except Figure 6 suggest adjusting the font size of the text to make it clearer for the audience.

References

Chu, Y., Zhang, H., Li, X., and Henneberger, J.: Exploring the effect of training set size and number of categories on ice crystal classification through a contrastive semi-supervised learning algorithm, Atmospheric Measurement Techniques, 18, 2781-2801, 2025.

Lindqvist, H., Muinonen, K., Nousiainen, T., Um, J., McFarquhar, G. M., Haapanala, P., Makkonen, R., and Hakkarainen, H.: Ice-cloud particle habit classification using principal components, Journal of Geophysical Research: Atmospheres, 117, 2012.

Praz, C., Ding, S., McFarquhar, G., and Berne, A.: A versatile method for ice particle habit classification using airborne imaging probe data, Journal of Geophysical Research: Atmospheres, 123, 13-472, 2018.

Schmitt, C. G., Järvinen, E., Schnaiter, M., Vas, D., Hart, L., Wong, T., and Stuefer, M.: Classification of ice particle shapes using machine learning on forward light scattering images, Artificial Intelligence for the Earth Systems, 3, 230 091, 2024.

Xiao, H., Zhang, F., He, Q., Liu, P., Yan, F, Miao, L., and Yang, Z.: Classification of ice crystal habits observed from airborne Cloud Particle Imager by deep transfer learning, Earth and Space Science, 6, 1877-1886, 2019.

Zhang, H., Li, X., Ramelli, F., David, R. O., Pasquier, J., and Henneberger, J.: IceDetectNet: A rotated object detection algorithm for classifying components of aggregated ice crystals with a multi-label classification scheme, Atmospheric Measurement Techniques, 17, 7109-7128. 2024.