

Interactive comment on “Unsupervised classification of snowflake images using a generative adversarial network and K -medoids classification” by Jussi Leinonen and Alexis Berne

Anonymous Referee #2

Received and published: 30 March 2020

Title: Unsupervised classification of snowflake images using a generative adversarial network and K-medoids classification

Authors: Leinonen and Berne

A new snowflake classification method using unsupervised machine learning methods is described in this study. It uses approximately two million snowflake images obtained with the Multi-angle snowflake camera (MASC) at various observational sites. This type of method developed will probably be used more and more by the scientific community producing enormous amount of precipitation particle photos using the MASC or other future technology. The paper is well and clearly written. The goal is stated clearly

C1

and the methodology is described in detail. It is, however, difficult for me to evaluate the choice of algorithms used. I only have a few minor comments divided into main comments and specific comments. They are listed below.

Main comments

1. Section 4: The unsupervised methodology developed to analyse the snowflake photos uses the K-medoids method instead of the K-means. These two techniques are described in section 3.4 but the author decided to use only one of them. I was wondering why the K-means is even described in the manuscript. It may be more straightforward to have only a paragraph describing the advantages using K-medoids methods with respect to the K-means and only describe the one used in the developed unsupervised classification.

2. Section 5: The authors describe the methodology used to classify the snowflakes using many K categories. In section 5.2.2, it shows that using 16 classes is more advantageous than using only 6. The authors demonstrate the feasibility and the quality of this unsupervised classification method. In section 5.2.4, it compares the unsupervised classification presented in section 5.2.2. It is difficult to see the link between sections 5.2.2 and 5.2.4 with section 5.2.3. Section 5.2.3 suggests that if we want to analysis microphysical properties of the snowflakes, one may want to have an expert doing it manually. Then, a list of different categories is given. This gives the impression that the method developed is not unsupervised while the goal of that section is probably simply to give an explanation of the limitation associated with the new unsupervised method developed. One suggestion would be to include section 5.2.3 into a discussion provided in the following section.

Specific comments

1. Line 33-35: It is mentioned that “snowflake imaging instruments have been actively developed in the recent years”. I am just curious to know if other instruments similar than the MASC exists.

C2

2. Verb tense: The authors should be consistent with the verb tense used in the manuscript. For example, line 122-124 should be past tense. Please verify throughout the manuscript to make sure that it is consistent.
3. Paragraph starting line 270: verb tense please double check.
4. Line 279-281: "However, we found that the latent. . . classification" should be clarified.
5. Line 283-297: please double check verb tense.
6. Paragraph starting line 323: please double check verb tense.
7. Figures 6, 7 and 9: Does the color code represent the same variable as the number in each square? I think that it is useful to have both the number on each square and the colorbar should be clarified in the figure caption.
8. Section 5.2.3 should probably be included in a following section comparing the unsupervised with the supervised methods.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2019-440, 2019.