Machine Learning Classification of Snowflakes to Enhance Microphysical and Scattering Characterization of Snow

Hein Thant(1), Mikhail Zhizhin (2), and Branislav M. Notaros(1)

(1) Colorado State University, Electrical & Computer Engineering Dept., Fort Collins, CO, USA
(hthant@colostate.edu, notaros@colostate.edu)

(2) Colorado School of Mines, Payne Institute for Public Policy, Golden, CO, USA (mikhail.zhizhin@gmail.com)

Abstract—**No two snowflakes are alike. However, they can be classified into over a hundred different categories based on their geometry. Recent developments in machine learning algorithms have led to the possibility of automatically classifying more classes of snowflakes more accurately and efficiently. We present supervised and unsupervised learning approaches to snowflake classification coupled with dimensional reduction techniques. Proper automatic and detailed classification of ice and snow hydrometeors enables advanced characterization of geometrical, microphysical, and scattering properties of particles, which, in turn, is essential for development of radar-based quantitative precipitation estimation of snow.**

I. Introduction

Classification of snowflakes has been recently invoked as an important factor in enriching our understanding of polarimetric radar signatures of snow, ice cloud processes and resulting precipitation. However, classification of over a hundred shapes of snowflakes [1] is a challenging problem for both atmospheric scientists and machine learning engineers due to the complex nature of snow. In the past decade, new multi-camera-based systems such as the Snowflake Measurement and Analysis System (SMAS) and Multi-Angle Snowflake Camera (MASC) were introduced. The images from these instruments are perfect to serve as input data for our models. Fundamentally understanding the data and extracting features is important to classify snowflakes. Convolutional Neural Networks (CNNs), while being a popular technique to classify images, fail to describe the features it learns from [2] and therefore make feature extraction and studying important.

II. Data Sources and Methodologies

The SMAS, a seven-camera system, and the MASC, a three-camera (modified to five), both capable of 3D reconstruction, particle size and fall speed measurement, serves as our primary source of data (Fig. 1). However, with the limited number of images that is already preprocessed and labeled, it is difficult for supervised learning. Our data includes the collected locations in Table 1. The methods used in this study include a simple CNN (Fig. 2) and additional dimensional reduction/feature extraction techniques using:

- PCA - Principal Component Analysis, dimensional reduction method that fits a p-dimensional ellipsoid.
  - Single Image PCA - fitting PCA over each 300x300 pixels image.
  - Tabular Data PCA - transforming all image into 90000-dimensional tabular data then fitting PCA.
- Hu Invariant Moments - Shape descriptors that give 7 unique moments of the silhouettes of images which are invariant over scaling, rotations or translations. This feature extraction method also disregards image texture.
- UMAP [3] - Uniform Manifold Approximation Projection, State-of-the-Art feature extraction technique that relies on different metrics such as Euclidean, Minkowski, etc. to cluster features and has supervised learning options.

Figure 1. Snowflake Measurement and Analysis System (left) and Multi-Angle Snowflake Camera (right).

TABLE I. Data Sources.

Data Source	MASC	SMAS	Location
MASCRAD	Y		CO, USA
ICE-POP	Y		Pyeonchang, KR
NASA-GPM-WFF	Y	Y	VA, USA
IMPACTS-UConn	Y	Y	CT, USA
MASCDB [4]	Y		Jura Mountains, CHE

III. Results and Analysis

We have included four classes of snowflakes for this preliminary study (Fig. 2): Aggregate, Small/Germ, Planar, Graupel-like. Fig. 3 shows the result from our CNN is positive for classifying Planes and germs but not for Aggregates and Graupel-like. This could be improved with a complex network.

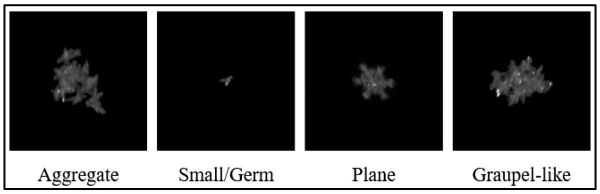

Figure 2. Snowflake Classes used for study.

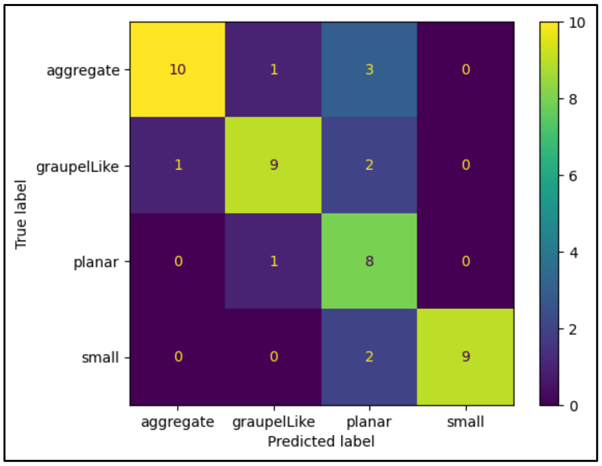

Figure 3. Confusion Matrix for Simple CNN architecture.

PCA on single image gives positive results for isolating small particles while the x axis provides a small amount of distinction symmetry between aggregates and planes (Fig. 4). Hu Invariant Moments could distinguish between shapes based on the silhouette (Fig. 5). However, since Graupel-like snowflake silhouettes are indistinguishable from enlarged small particles, their differences could not be differentiated. UMAP with Euclidean [5] on feature extraction has general clustering between classes but the clusters overlap (Fig. 6). It shows separation between sizes on the y-axis and a general decrease in symmetry on the diagonal axis. This presents similar features to PCA.

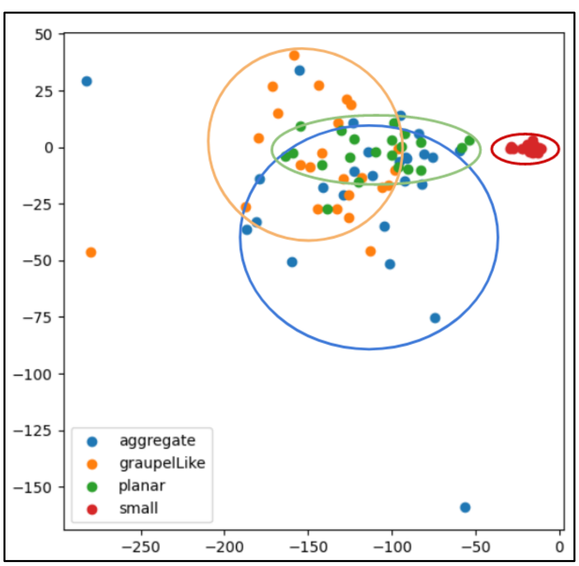

Figure 4. PCA Single Image.

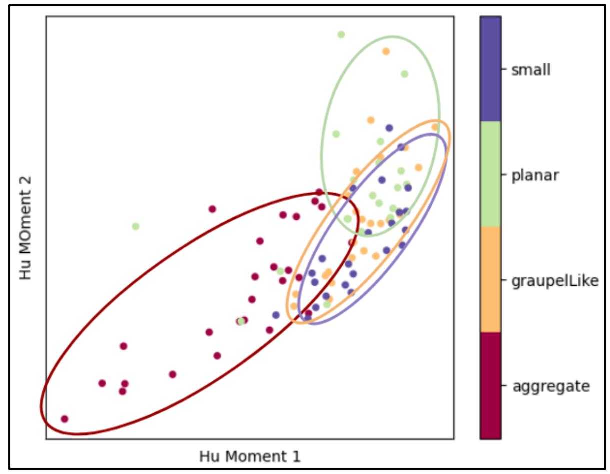

Figure 5. Hu Invariant Moments.

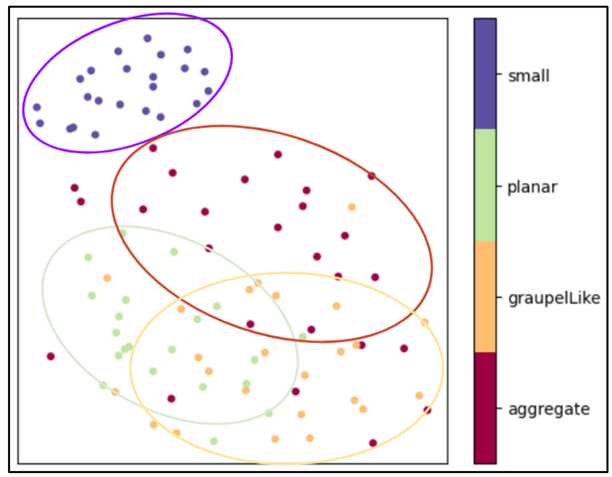

Figure 6. UMAP Supervised.

ACKNOWLEDGMENT

This work was supported by the National Science Foundation under grant AGS-2029806.

REFERENCES

[1] K. Kikuchi, T. Kameda, K. Higuchi, and A. Yamashita, "A global classification of Snow Crystals, ice crystals, and solid precipitation based on observations from middle latitudes to polar regions," *Atmospheric Research*, vol. 132-133, pp. 460–472, 2013.

[2] A. Hicks and B. M. Notaroš, "Method for classification of snowflakes based on images by a Multi-Angle Snowflake Camera using convolutional neural networks," *J. Atmos. Oceanic Technol.*, vol. 36, pp. 2267-2282, 2019.

[3] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: Uniform manifold approximation and projection," *Journal of Open Source Software*, vol. 3, no. 29, p. 861, 2018.

[4] J. Grazioli, G. Ghiggi, A.-C. Billault-Roux, et al., "MASCDB: A database of images, descriptors, and microphysical properties of individual snowflakes in free fall," *Sci Data*, vol. 9, p. 186, 2022.