

Reducing the Learning Population Size for Morphological Classification of Natural Materials: Example of Gold Grains from Backscattered Electron Images

Alexandre Néron¹, Réjean Girard¹, Jonathan Tremblay¹, L. Paul Bédard² and Kevin Bouchard³

¹. IOS Services Géoscientifiques Inc., Saguenay, Canada.

². LabMaTer, Sciences de la Terre, Université du Québec à Chicoutimi, Saguenay, Canada.

³. Université du Québec à Chicoutimi, Computer Department, Saguenay, Canada.

Morphological classifications of natural materials are used in a wide variety of applications. Visual classification of some natural materials such as minerals, characterized by their wide morphological variations, are challenging. In some instances, differences between categories are subtle or forming continuum, causing high misclassification rates. Supervised machine learning (ML) is a powerful tool to automate and improve classification. However, most image classification algorithms require a large labeled dataset in which the quality of the initial labeling is critical. Since visual labeling of hundreds of thousands of images is a tedious and biased process, a specific learning strategy is proposed to circumvent such issues. This paper describes the classification procedure for gold grain morphologies obtained from backscattered electron images.

Determining the morphology of gold grains is routinely used in gold deposit exploration in glacial tills. Gold grains are eroded and scattered from their source rocks by the movement of glaciers. Due to its malleability, shapes and surface characteristics such grains contains crucial information on their geological setting and transport history. However, an accurate classification is a prerequisite in order to ensure realistic interpretation. Three shapes were first proposed [1] assumed to reflect the distance of transport, from “pristine” to “modified” to “reshaped”. However, the morphology and surface characteristics of very minute gold grains (< 30 μm) are hardly perceptible by optical microscopy commonly used for such. Moreover, this method is highly user-dependent. To circumvent those limitations, the use of an automated SEM-based routine was developed to detect and acquire a high magnification BSE image of gold grains from heavy mineral concentrates. This method greatly increases the resolution of surface characteristics and morphology of gold grains. Since minute gold grains have a lower probability of being deformed during transport, the likelihood of their original morphology preservation is higher than larger ones (> 30 μm). To better understand their geological settings, the definition of “pristine” ought to be further subdivided into four morphological classes based imprints related to their relation to adjacent mineral (Fig. 1): A) “Remobilized” grains are interpreted as crack-filling flakes, which are very fragile due to their minimal thickness; B) “Crystalline” grains crystallized according to their cubic crystal system. Their edges are sharp with constant face angles and show minimal signs of deformation. C) “Mutual” grains are those who co-crystallized with minerals of similar crystallization habits. Their edges are curvy and convex. D) “Intergranular” grains are those with discernible sharp edges, but with concave faces, suggesting gold filled the boundaries of host minerals.

Since early 2016, a total of 8 591 BSE images of gold grains were sorted by a trained geologist. Multiple automated classification algorithms have been tested on those images, with yield a maximal classification success of 64% when used upon the validation population. Such classification rate is four times higher than a random classifier, but still sufficient for decisional application. Variations of accuracy between different classification algorithms are not significant, suggesting issues from the learning dataset. The proposed strategy to improve the learning dataset is divided in three steps.

- 1) Subsets of the learning dataset were generated using the 100 best candidates in each category. The selection, based on preservation of discriminating features, was performed by trained geologist and validated by two others.
- 2) The resulting subsets were used as training sets by the classification algorithm. Transfer learning, performing properly while using small datasets [2], was applied on the Google's Convolutional Neural Network (CNN) architecture Inception-V3 [3] using TensorFlow [4]. In a high-level perspective, a CNN takes an image as the input, finds discriminating features and, on the last layer, classifies the images based on their features. Transfer learning on a CNN keeps the lower layers and re-trains only the last layer with new categories. The accuracy of the learning on the best candidate subset reached 99,3% on the validation set (25% of the dataset), meaning that categories are unambiguous enough to be distinguished with ML and that the chosen model has a similar feature space than our training images.
- 3) Morphologies of grains from the initial dataset were classified using algorithms from step 2. Predictions were validated by a trained geologist and globally rechecked by another professional. During the validation process, the geologist had to decide if the grains were correctly classified, and if not, manually reclassified. The accuracy of the machine classification was proofed at 87%. The reclassification by the geologist in step 3 is kept in preference over the machine prediction model.

Improving the classification algorithm is performed by iterating the learning process presented in step 2 on the entire classified population. A final accuracy of 92% is achieved, which is adequate for this type of application. Compared to the initial classification, this suggests that 49.7% of grains were misclassified using visual interpretation, reinforcing the need of automated classification.

References:

- [1] DiLabio, R.N.W., Geological Survey of Canada, **90-1C** (1990), p. 323-329.
- [2] Pan, S.J., and Yang, Q., IEEE Transactions on knowledge and data engineering **22.10** (2010), p. 1345-1359.
- [3] Szegedy, C. *et al.*, IEEE Conference on Computer Vision and Pattern Recognition (2016).
- [4] Abadi, M. *et al.*, OSDI **16** (2016).
- [5] The authors acknowledge funding from Fonds de recherche Québec Nature et Technologies (Projet de recherche orientée en partenariat) to LPB. Vincent Tao, Sanmei Gao and Phillippe Pagé, are thanked for their contributions to this work.

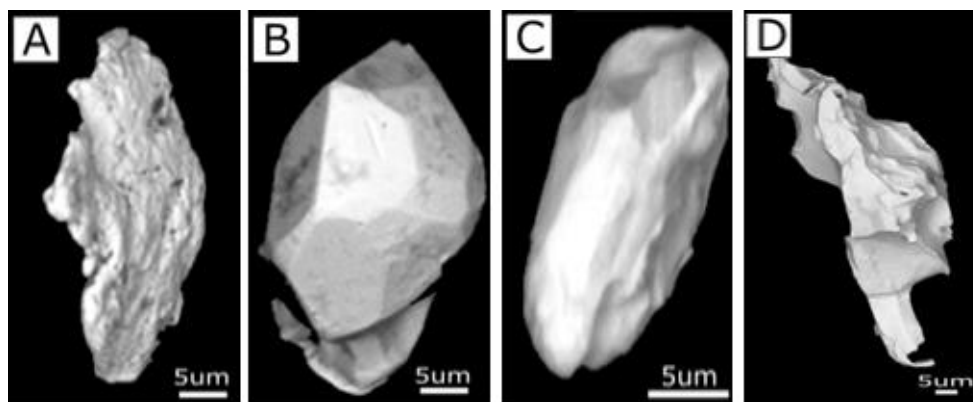


Figure 1. Morphologies of pristine grains. A) Remobilized, B) Crystalline, C) Mutual D) Intergranular.