Introduction

- High demand for cyber-infrastructure which enables geological image data to be shared.
- Develop a system which accurately identifies whether an input image contains a sigma-clast, even with limited data.

Dataset

- Allows for smaller datasets to become larger, as long as data is not augmented to a different classification (e.g., Distorting Sigma-Clast image to become more of a Non-Sigma-Clast)

Data Augmentation

- Effects Utilized:
  - Rotation range: 40°
  - Height/width shift: 10%
  - Shear range: 0.2 rad ccw
  - Zoom: 30%
  - Horizontal flipping: true
  - Fill mode: “reflect”

Transfer Learning and Fine Tuning

- Allow a Convolutional Neural Network to correctly learn over a very large dataset
- Take the Convolutional Neural Networks and utilize its features instead of creating our own.
- Fine tuning modifies the last few layers to ensure we look for the complex structures we want.

Neural Network Ensembles

- Takes Multiple Convolutional Neural Networks and combines their output.
- This method ensures that we reduce the variance between multiple models by taking the average of each model to compute the classifications.

Artificially Increasing the Dataset with Oversampling

- We can artificially increase the largest class of data we have, then use oversampling to create copies of the smaller dataset to match the dataset.

Full Stack Architecture for Web-based Classification

- Front-End: React JS
  - Handles the user interface, sets necessary objects to be sent to the backend.
- HTTP Client: Axios
  - “Connects” the front-end to the back-end using HTTP.
- Microframework: Flask
  - Designed for Python; handles the HTTP requests that come from Axios.
- Back-End: Python
  - Creates functionality to resize and classify images
- Classification: Keras and Tensorflow
  - The libraries that allow for the prediction and classification of images.

Testing Data Augmentations

- Summary:
  - F1 score is an evaluation metric that is a trade-off between the recall (true positive rate) and precision (fraction of relevant classifications).

Conclusion and Future Research

- Significant improvement of accuracy can be seen among the system of networks.
- Being able to classify the rotation of a Sigma-Clast (i.e., Counter-Clockwise/Clockwise)
- Allow for K-Fold cross validation evaluation
- Generalize the model for multiclass classification.
- Web Application
  - Enhance UI – color scheme, fill whitespace effectively
  - Allow for prediction options such as ensembles and fine-tuning
  - Host application on a live site

Acknowledgements

Special thanks to Earthcube NSF Grant for funding this research, as well as Dr. Gurman Gill for guiding the team.