EGU23-2897, updated on 30 Nov 2023
https://doi.org/10.5194/egusphere-egu23-2897
EGU General Assembly 2023

Automatic Classification of THEMIS All-Sky Images via Self-Supervised Semi-Supervised Learning

Jeremiah Johnson[1], **Dogacan Ozturk**[2], Hyunju Connor[3], Donald Hampton[2], Matthew Blandin[2], and Amy Keesee[4,5]

[1]University of New Hampshire, Applied Engineering and Sciences, United States of America (jeremiah.johnson@unh.edu)
[2]The Geophysical Institute, University of Alaska-Fairbanks, United States of America
[3]NASA Goddard Space Flight Center, Greenbelt, MD, United States of America
[4]Department of Physics, University of New Hampshire, United States of America
[5]The Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, United States of America

Dynamic interactions between the solar wind and the magnetosphere give rise to dramatic auroral forms that have been instrumental in the ground-based study of magnetospheric dynamics. The general mechanism of aurora types and their large-scale patterns are well-known, but the morphology of small- to meso-scale auroral forms observed in all-sky imagers and their relation to magnetospheric dynamics and the coupling of the magnetosphere to the upper atmosphere remain in question. Machine learning has the potential to provide answers to these questions, but most existing auroral image data lack the ground-truth labels required for supervised learning and conventional statistical analyses. To mitigate this issue, we propose a novel self-supervised semi-supervised algorithm to automatically label the THEMIS all-sky image database. Specifically, we adapt the self-supervised Simple framework for Contrastive Learning of Representations (SimCLR) algorithm to learn latent representations of THEMIS all-sky images. These representations are finetuned using a small set of manually labeled data from the Oslo Aurora THEMIS (OATH) dataset, after which semi-supervised classification is used to train a classifier, beginning by training on the manually labeled OATH dataset and gradually incorporating the classifier's most confident predictions on unlabeled data into the training dataset as ground-truth. We demonstrate that (a) classifiers fit to the learned representations of the manually labeled images achieve state–of–the–art performance, improving the classification accuracy by almost 10% over the current benchmark on labeled data; and (b) our model's learned representations naturally cluster into more clusters than manually assigned categories, suggesting that existing categorizations are coarse and may obscure important connections between auroral types and their drivers. Finally, we introduce AuroraClick, a citizen science project with the goal of manually annotating a large representative sample of THEMIS all-sky images for the validation of our current models and the training of future models.