Project Details

ROSES ID: NNH21ZDA001N-LWSTM
Selection Year: 2021
Program Element: Data, Tools, & Methods

Project Title:
Producing Homogeneous, Machine-Learning Ready Auroral Image Databases Using Unsupervised Learning

PI Name: Jeremiah William Johnson
PI Email: jeremiah.johnson@unh.edu
Affiliation: University of New Hampshire, Durham
Project Member(s):
  - Hampton, Donald L;Collaborator;University Of Alaska, Fairbanks
  - Connor, Hyunju K;Collaborator;University Of Alaska, Fairbanks
  - Ozturk, Dogacan Su;Co-I/Institutional PI;University Of Alaska, Fairbanks

Summary:
Dynamic interactions between solar wind and magnetosphere gives rise to dramatic auroral forms that have been instrumental in the ground-based study of magnetospheric dynamics. Although the general mechanism of aurora types and their large--scale patterns are well-known (Newell et. al. 2009), the morphology of small- to meso-scale auroral forms observed in all-sky imagers and their relation to the magnetospheric dynamics are still in question. A better understanding of the morphology of auroral forms is critical to our understanding of magnetospheric dynamics and the coupling of the magnetosphere to the upper atmosphere. Machine learning offers the possibility of surfacing new knowledge in this area, but most existing auroral image databases are not yet machine learning-ready. A key issue is the lack of ground-truth labels: most widely-used machine learning algorithms are supervised, and without ground-truth labels, cannot be used.

The scientific goal of this project is to deliver a large-scale, homogeneous, machine learning-ready database of auroral images that will enable machine learning-driven investigations into the morphology of auroral forms and the corresponding relationship between these forms and magnetospheric dynamics.

To achieve this goal, we will develop a state-of-the-art unsupervised machine learning algorithm capable of automatically labeling 16 years of white light auroral image data from Time History of Events and Macroscale Interactions during Substorms (THEMIS) ground-based All-Sky Imager (ASI) array. Our approach avoids the necessity of ground-truth labels during training by learning latent representations of the input data that capture inherent structural relationships. These representations can then be used for downstream tasks such as automatic labeling and cluster analysis or investigated in their own right. This work will produce the largest publicly-available, labeled, homogeneous, machine-learning ready auroral image database created to date. It will enable the space science community to conduct statistical studies on the relationship between different categories of auroral images, near-earth solar wind conditions, and geomagnetic disturbances at the earth's surface that were not previously possible. It is relevant to the high-level goal to "Determine the dynamics and coupling of Earth's magnetosphere, ionosphere, and atmosphere and their response to solar and terrestrial input" from the Heliophysics Decadal Survey. The machine learning-ready dataset produced by this research, along with the models and software necessary to reproduce the results, will be delivered to the Space Physics Data Facility on or before the conclusion of the project on October 31, 2023.

Publication References:

no references