JGR Space Physics

METHOD

10.1029/2023JA032096

Automatic Encoding of Unlabeled Two Dimensional Data Enabling Similarity Searches: Electron Diffusion Regions and Auroral Arcs

A. W. Smith[1], I. J. Rae[1], J. E. Stawarz[1], W. J. Sun[2,3], S. Bentley[1], and A. Koul[4]

[1]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle upon Tyne, UK, [2]Space Sciences Laboratory, University of California, Berkeley, CA, USA, [3]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA, [4]Pinterest Inc, San Francisco, CA, USA

Abstract Critically important phenomena in Earth's magnetosphere often occur briefly, or in small spatial regions. These processes are sampled with orbiting spacecraft or by fixed ground observatories and so rarely appear in data. Identifying such intervals can be an incredibly time consuming task. We apply a novel, powerful method by which two dimensional data can be automatically processed and embeddings created that contain key features of the data. The distance between embedding vectors serves as a measure of similarity. We apply the state-of-the-art method to two example datasets: MMS electron velocity distributions and auroral all sky images. We show that the technique creates embeddings that group together visually similar observations. When provided with novel example images the method correctly identifies similar intervals: when provided with an electron distribution sampled during an encounter with an electron diffusion region the method recovers similar distributions obtained during two other known diffusion region encounters. Similarly, when provided with an interesting auroral structure the method highlights the same structure observed from an adjacent location and at other close time intervals. The method promises to be a useful tool to expand interesting case studies to multiple events, without requiring manual data labeling. Further, the models could be fine-tuned with relatively small set of labeled example data to perform tasks such as classification. The embeddings can also be used as input to deep learning models, providing a key intermediary step—capturing the key features within the data.

Plain Language Summary Space plasma physics missions and observatories often collect large datasets. These data do not come with interpretation as to what is happening at each time interval, this must be applied manually by experts. This is a very time consuming process and can inhibit authors from studying interesting processes. In this work we showcase a ground breaking technique that can process huge quantities of data automatically to produce encoded versions of the original data. By comparing the encoded versions we can find similar observations without needing to manually search through the data. The method works exceptionally with rare events, finding two of only a few examples in several thousand images.

Key Points:

- Method creates encoded representations from unlabeled two dimensional data, small intra-embedding distance can indicate similar examples
- When provided with an electron distribution from a known diffusion region encounter, the method successfully finds other known examples
- When given an all sky image of an auroral arc, the method locates an image of the same arc from an adjacent location

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. W. Smith,
andy.w.smith@northumbria.ac.uk

Citation:

Smith, A. W., Rae, I. J., Stawarz, J. E., Sun, W. J., Bentley, S., & Koul, A. (2024). Automatic encoding of unlabeled two dimensional data enabling similarity searches: Electron diffusion regions and auroral arcs. *Journal of Geophysical Research: Space Physics*, *129*, e2023JA032096. https://doi.org/10.1029/2023JA032096

Received 18 SEP 2023
Accepted 16 JAN 2024

©2024. The Authors.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The dynamics of the near-Earth space environment are often dictated by rare or small-scale processes. Observations of such phenomena constitute a very small portion of the large, high time resolution datasets that are collected by both orbital platforms and ground based observatories. These large datasets—by their very nature—have no labels/interpretation, save for those applied by experts during time consuming exercises.

For example, close to the site of magnetic reconnection the plasma becomes decoupled from the local magnetic field, forming diffusion regions (Burch, Torbert, et al., 2016; Øieroset et al., 2001). Fortuitous spacecraft encounters with such regions often number in the low tens of examples (e.g., Eastwood et al., 2010; Lenouvel et al., 2021; Rogers et al., 2019, 2023). Nonetheless through these rare identifications we seek to understand the fundamental nature of these processes. Broadly, studies into such phenomena can be grouped into two main categories: case studies (e.g., Burch, Torbert, et al., 2016; Øieroset et al., 2001; ; Phan et al., 2018) and statistical assessments (e.g., Eastwood et al., 2010; Rogers et al., 2019, 2023). The problem lies in extrapolating from interesting case studies to a robust, comprehensive statistical sample. Manual expert selection may be thorough, given sufficient time, however it runs the risk of expert judgment varying over the period of the survey. A key limitation here is the availability of expert time, even with the adoption of advanced, time saving tools (e.g., Louis et al., 2022).

In recent years automatic identification schemes have become more popular, leveraging increased computational power to reduce the expert time required to query large datasets. Methods have been designed for tasks such as identifying encounters with magnetic structures in time series data from spacecraft, either from an empirical base (e.g., Smith et al., 2017) or using more advanced, but black box methods (e.g., Garton et al., 2021). Empirical methods often aim to identify particular signatures in one dimensional parameters such as changes or peaks in the observed magnetic field (e.g., Bergstedt et al., 2020; Huang, Zhao, et al., 2018; Sarkango et al., 2022), or when thresholds are exceeded (e.g., plasma bulk flow velocity, wave power) (e.g., Hamrin et al., 2015; Smith et al., 2020). Unfortunately these one-dimensional methods cannot often be generalized to more complex, higher dimensional datasets. Two dimensional data, such as particle distributions or images, can have single numerical parameters features extracted (e.g., Lenouvel et al., 2021), which may then be processed in a similar fashion to other one dimensional parameters. While this manual dimensionality reduction can be carefully designed to target specific properties, it can be laborious to design a robust method. Most often, two dimensional data are processed as images, for example, convolutional neural networks have been used to classify auroral images (e.g., Clausen & Nickisch, 2018; Endo & Matsumoto, 2022; Shang et al., 2023; Zhong, Ye, et al., 2020). However, we note that all of these methods require manually identified—"labeled"—datasets on which to train.

Unsupervised techniques, those that do not require manual labeling, have become more popular for use with higher dimensional datasets in the last few years. However, such methods are mostly used to group ("cluster") similar observations, for example, to identify observations in distinct magnetospheric regions based on plasma moments (e.g., Bloch et al., 2020; Innocenti et al., 2021) or distributions (e.g., Bakrania et al., 2020; Killey et al., 2023). For two dimensional data, techniques such as auto-encoders may be used to reduce the dimensionality of the data. This allows similar observations to be grouped together in the smaller dimensional space—a space that is assumed to be representative of the original data (e.g., Bakrania et al., 2020; Killey et al., 2023). However, typically such methods directly use the pixel values in the data and thus care must be taken if a key feature of the image is known to move within an image or vary in brightness.

Here we present the application of a novel solution to enable the leap from case studies to statistical surveys without manual labeling, allowing the use of high dimensional datasets without requiring a priori dimension reduction or the development of manual thresholds. The method does not rely on precise pixel values, but rather extracts "features" from the images. We show the method applied to two distinct datasets, in situ electron distributions and auroral all sky images.

2. Data and Model

2.1. Model Architecture

Most commonly, two dimensional data (i.e., images) are processed through supervised methodologies, where images are linked to labels. As an intermediate step, these methods produce encoded versions of the images (embeddings), containing the critical descriptive information (i.e., features of the image) without extraneous details specific to the image (i.e., specific pixel values). These descriptive encodings are ultimately what is then linked to the label of interest. However, recent advances allow the creation of encodings without the need for labels.

In this work we demonstrate a model based upon the SimSIAM method using self-supervised learning (Chen & He, 2021). The main aim is to generate encoded representations (i.e., embeddings) that could be passed to downstream tasks (e.g., classification or as input to other models). Critically, the chosen method does not require labeled data and instead learns to extract the key features of the provided images using data augmentation. Often a "pretext" supervised downstream task is used in this process: the model is trained to perform a task such as image classification, which aids the development of the encoder. This may be performed as an additional step, refining the embedding process after a period of unsupervised training, which then requires only a small subset of the data to be labeled. In these cases the model performance has matched that obtained by models trained using more typical supervised techniques (Chen & He, 2021). However, for our use case—with unlabeled space plasma physics data—we will show that we can generate embeddings of a sufficient quality for our purpose (finding specific, rare features) without a pretext task. Self-supervised learning has been shown to work particularly well with imbalanced datasets (Liu et al., 2022), as are common in space plasma physics, and has been used to great success with data such as satellite imagery (e.g., Agastya et al., 2021; Alosaimi et al., 2023) and astronomical images (e.g., Hayat et al., 2021).

Most common machine learning methods, particularly supervised techniques, optimize the "loss" between the model output and a target during the training process, L2 loss for example. However, self-supervised learning uses a contrastive loss function. In place of a fixed target, the loss is minimized between two different representations generated by the model. In a typical contrastive loss function, the loss will be low if two representations (embeddings, or feature vectors) are similar, and large if they are distinct. Contrastive loss functions may also measure that the representation is distinct from those generated from independent data. We direct the interested reader to an excellent summary contained within He et al. (2019). A number of different architectures have been proposed, including MoCo (He et al., 2019), SimCLR (Chen, Fan, et al., 2020), MoCo V2 (Chen, Fan, et al., 2020), BYOL (Grill et al., 2020),and SimSIAM (Chen & He, 2021), with subsequent developments tending to increase performance while reducing limitations or training restrictions.

Figure 1 shows an outline of the selected SimSIAM-based model design. Here, each image in the training set ("X") is processed into two augmented versions ("X_1" and "X_2"). It is from these augmentations that the model learns the key features of the images, that is, those that distinguish each image. The augmented images are processed by the encoder network, E, consisting of a backbone and a projection feed-forward layer (e.g., Chen, Fan, et al., 2020). For our backbone we have chosen the ResNET architecture (He et al., 2015), in particular the ResNet-18 (18-layer) model, as we found it to strike a good balance between model complexity (and therefore computational cost) and performance. The encoded representation for one augmented image is then transformed by a prediction head, H, matched to the other encoded representation, and the contrastive loss evaluated. As in SimSIAM models, the negative cosine similarity is chosen as the loss function, and is minimized between the two output vectors during training. The inclusion of a stop gradient in the model has been found to prevent the collapse of the training process (i.e., the outputs collapsing to a constant vector representation) (Chen & He, 2021).

Simply put, the model attempts to maximize the similarity between the two augmented views of the same image. The method trains such that the embeddings generated by the encoder are invariant to the chosen augmentations.

In this work we considered two popular approaches: the SimCLR and SimSIAM architectures. The SimCLR architecture is similar to the chosen SimSIAM method, however it uses adversarial comparisons: it requires "negative" sample pairs: the dissimilarity between distinct images (and their augmentations) is maximized during training (Chen, Fan, et al., 2020). While the SimCLR architecture has been used successfully to classify auroral images (Johnson et al., 2021), we found the high degree of auto-correlation between unlabeled images from time series made the training process challenging.

2.2. Example Use Cases: Motivation and Data

2.2.1. Case A: MMS Electron Distributions

Our first example dataset concerns in situ data, specifically electron distributions. Magnetic reconnection is a powerful energy conversion process in magnetized plasmas that involves the reconfiguration of the magnetic field. Close to the site of reconnection the plasma decouples from the magnetic field, forming "diffusion regions" (Øieroset et al., 2001). Due to their smaller mass, the electron diffusion region occupies a smaller region within the broader ion diffusion region. Nonetheless, electron diffusion regions have been identified at both the magnetopause (e.g., Burch, Torbert, et al., 2016; Chen et al., 2017; Graham et al., 2017; Lenouvel et al., 2021; Torbert et al., 2016, 2017; Zhong, Ye, et al., 2020), and within the magnetotail of the Earth (e.g., Torbert et al., 2018; Chen et al., 2019; Farrugia et al., 2021; Hasegawa et al., 2022; Huang, Jiang, et al., 2018; Li et al., 2019; Li et al., 2023; Zhou et al., 2019). The demagnetized electrons within the electron diffusion region are deflected and accelerated by electric fields resulting in a distinctive "crescent" shape in the observed velocity distribution (e.g., Burch, Torbert, et al., 2016). However, while these crescents are symptomatic of an electron diffusion region they may also be observed in other locations where there are electron-scale boundaries, such as within the foreshock (Gao et al., 2023).

We use data from the magnetospheric multiscale (MMS) Mission (Burch, Moore, et al., 2016), in particular fast mode data (4.5 s cadence) from the fast plasma investigation (FPI) (Pollock et al., 2016). Typically, crescents in electron distributions are identified in electron burst mode (33 Hz) data (e.g., Burch, Torbert, et al., 2016), however for this work we have chosen the lower cadence, fast mode data due to its greater coverage, and completeness (i.e., fewer data gaps). We are interested in the electron distribution relative to the magnetic field, in particular the plane perpendicular to the magnetic field, and we use PySPEDAS to

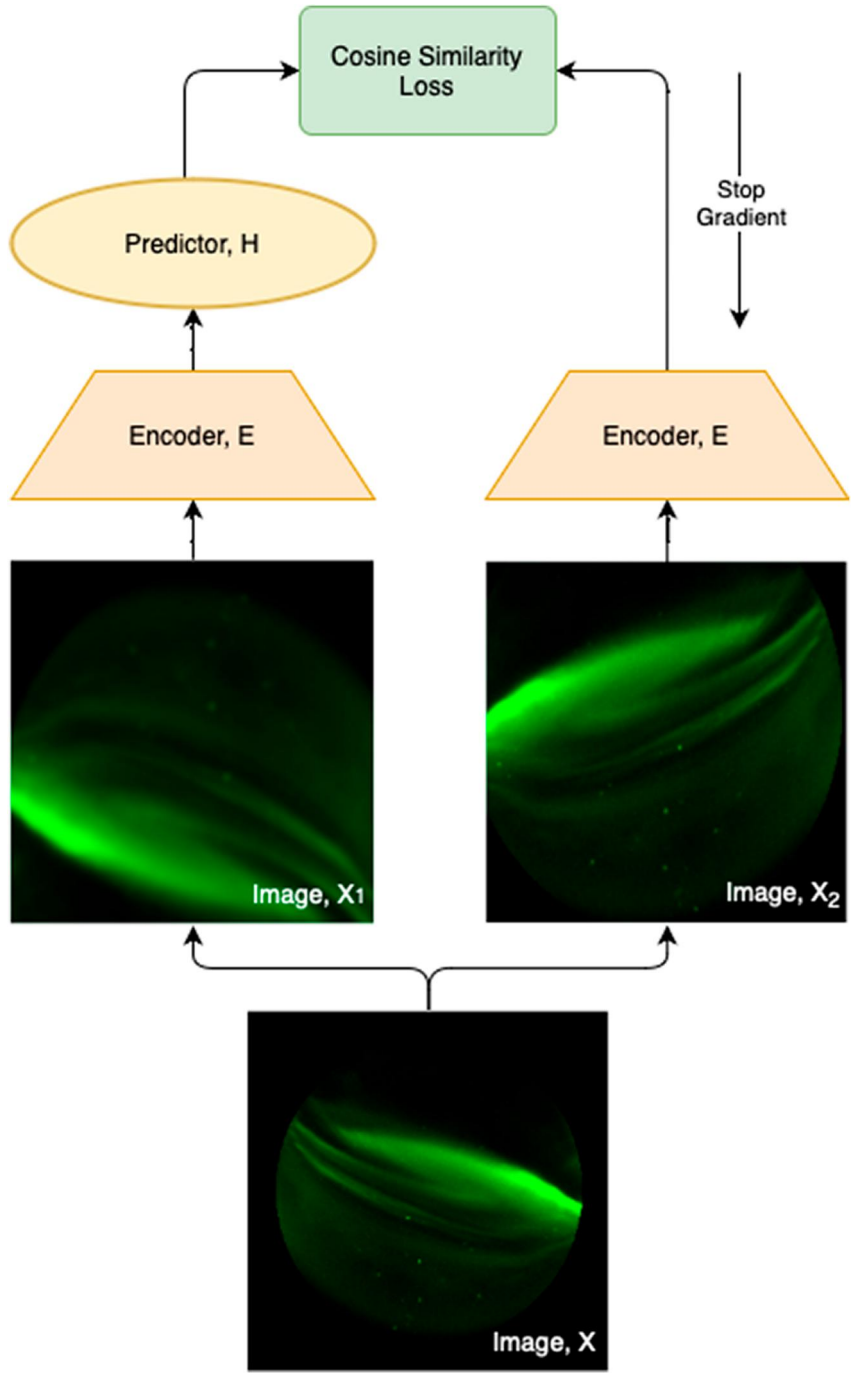


Figure 1. A schematic outlining the SimSIAM model architecture (Chen & He, 2021), as applied to an example THEMIS auroral all sky image (X) (Donovan et al., 2006). The image is processed into two augmented forms, X_1 and X_2. Each augmented image is then processed by an encoder network, E. One encoded representation is then processed by a prediction head, H, before the cosine similarity loss between the two representation is evaluated.

rotate the distributions into this plane (Angelopoulos et al., 2019). In particular we use the "perp" coordinate system where the x axis is V_{bulk} projected onto the plane normal the magnetic field and the y axis is $B \times V_{bulk}$. We take a 2D slice through the distribution, within 20° of the "perpendicular" plane. More detail about the data preparation steps can be found in the Text S1 in Supporting Information S1, though we note here that the distributions are limited to electron energies between 2 and 30 keV in order to increase the prominence of the crescent feature in the resulting images. We use four, 2-h intervals around known electron

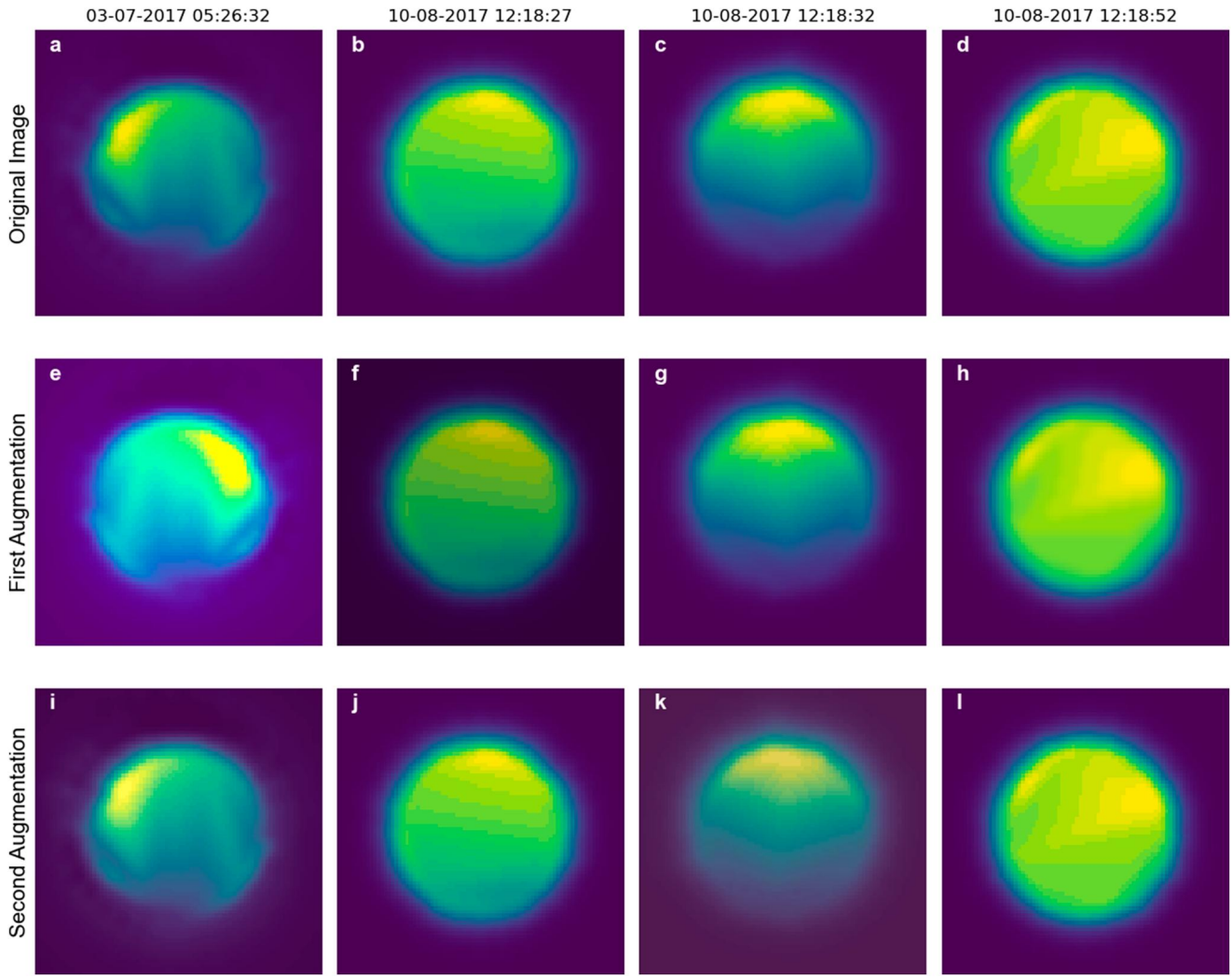


Figure 2. Four example images with augmented versions. Top (a–d): The original images. Middle (e–h): the first randomly augmented image (see Text S2 in Supporting Information S1). Bottom (i–l): the second random augmented image.

diffusion region encounters in the magnetotail by MMS1 as our dataset. In particular these intervals are on: 17th June 2017 (Farrugia et al., 2021; Huang, Jiang, et al., 2018), 3rd July 2017 (Chen et al., 2019); 11th July 2017 (Torbert et al., 2018); and 10th August 2017 (Zhou et al., 2019; Hasegawa et al., 2022; Li et al., 2019). In total this gives us a dataset of 5,740 images. We note that this dataset is smaller than is typically used for general self-supervised learning, however we consider that the variation between images is much smaller than in many applications (e.g., where there are thousands of different image categories).

The choice of augmentations, those that produce "X_1" and "X_2", are key to the training of the model. Augmentations that can be applied include: random rotations, mirroring (vertical and horizontal), and changes to color, for example, the hue, saturation and brightness. Our chosen augmentations were selected through an iterative process involving visual inspection and the results of test model training runs. Figure 2 shows four examples of the augmentation process. The original image is presented at the top (Figures 2a–2d), while two random augmentations of each image are presented below (augmentation one: Figures 2e–2h, augmentation two: Figures 2i–2l). It is clear for the first example in Figure 2a (3rd July 2017) that augmentation one has been flipped, the brightness increased and blur applied (Figure 2e), while the second augmentation (Figure 2i) is slightly darker but has had fewer augmentations applied. As described above, as the model trains it compares the encoded versions of the first and second augmentations, maximizing their similarity. In the Text S2 in Supporting Information S1 we detail the augmentations that were selected for both example datasets.

2.2.2. Case B: Auroral All Sky Images

Our second example application is to auroral all sky images. The ionosphere provides a surface on which activity in near-Earth space can be projected through the aurora. For decades it has been observed that during magnetospheric substorms the aurora is noted to brighten significantly and move in latitude (e.g., Akasofu, 1964). However, small-scale structure within the aurora can reveal details of the processes in near-Earth space (Forsyth et al., 2020; Henderson, 2013), important structures include auroral beads (e.g., Kalmoni et al., 2017; Ohtani & Motoba, 2022; Tian et al., 2022), streamers (e.g., Nishimura et al., 2011; Sergeev et al., 1999, 2004), and omega bands (e.g., Opgenoorth et al., 1983; Partamies et al., 2017; Wild et al., 2000). The availability and coverage of the datasets from ground based observatories—compared to in situ satellite observations—allows the near-continuous tracking of critical magnetospheric processes. Automated techniques have been developed to perform tasks such as the identification and tracking of auroral forms (Syrjäsuo & Donovan, 2002), enabling large scale surveys (Syrjäsuo & Donovan, 2004; Syrjasuo et al., 2007). Various auroral classification schemes have been used, ranging from simply "aurora" and "no aurora" (Syrjäsuo & Donovan, 2002), through to more complex categorizations including features such as "arc", "discrete" and "diffuse" auroral forms (e.g., Clausen & Nickisch, 2018; Endo & Matsumoto, 2022; Johnson et al., 2021; Shang et al., 2023; Zhong, Ye, et al., 2020). However, such categorizations rely on large, manually labeled datasets with empirically determined features (e.g., Clausen & Nickisch, 2018). Recent data-driven exploration has suggested that broad categories may be overly simplistic, potentially obscuring important physics (Johnson et al., 2021).

In this work we use a database of all sky images from the THEMIS cameras (Donovan et al., 2006), originally processed and labeled by Clausen and Nickisch (2018). Details of the data, image processing, augmentation and initial results are present in the Texts S2, S3, and S5 in Supporting Information S1.

2.3. Model Training

We now discuss the training procedure adopted for both example datasets. First, we withhold 30% of the data as a validation set, and monitor the negative cosine loss on both this and the training set, checking that the loss decrease is smooth and plateaus as expected. Further, we monitor the output vectors (embeddings) using a "collapse" metric to ensure that the vectors have not collapsed to a single output (i.e., Chen & He, 2021; Li et al., 2022). We perform a grid search (using Weights and Biases: https://wandb.ai) to determine the batch size, training epochs and the momentum of the optimizer. More details can be found in the Text S4 in Supporting Information S1. Performance and loss will vary depending upon the dataset and augmentations. Code to train the model is provided in (Smith, 2023).

3. Results

The end result of the training process is a model that will evaluate images, generating embedding vectors. In particular, we extract the trained ResNet-18 "backbone" from the encoder (E, in Figure 1), and apply it to our images to produce 512 dimensional embeddings.

3.1. Visualizing Embeddings

For visualization, we now reduce the 512 dimensional embeddings to a 2D space. In this work we use a t-SNE technique (Van Der Maaten & Hinton, 2008), a variation of the stochastic neighbor embedding (SNE) method (Hinton & Roweis, 2003). We note that this method is designed for visualizing high dimensional spaces, but has a cost function that is not convex: different runs will result in distinct visualizations. We then create a fixed grid over the two-dimensional space, selecting and presenting only those images closest to the center of each grid cell (to improve clarity), producing Figure 3. As this is a two-dimensional representation of a much larger dimensional space we will not comment quantitatively.

In Figure 3, showing the distribution of MMS electron distributions, we can see that the more isotropic distributions dominate the visualization, particularly in the top center. However, the distributions on the lower right show a range of anisotropies (i.e., asymmetries) of various different types. Importantly for this work, we can see that several of the distributions at the lower right of the visualization show evidence of crescent-like features above their center, as in the example in Figure 2c. During the data processing steps we deliberately did not remove

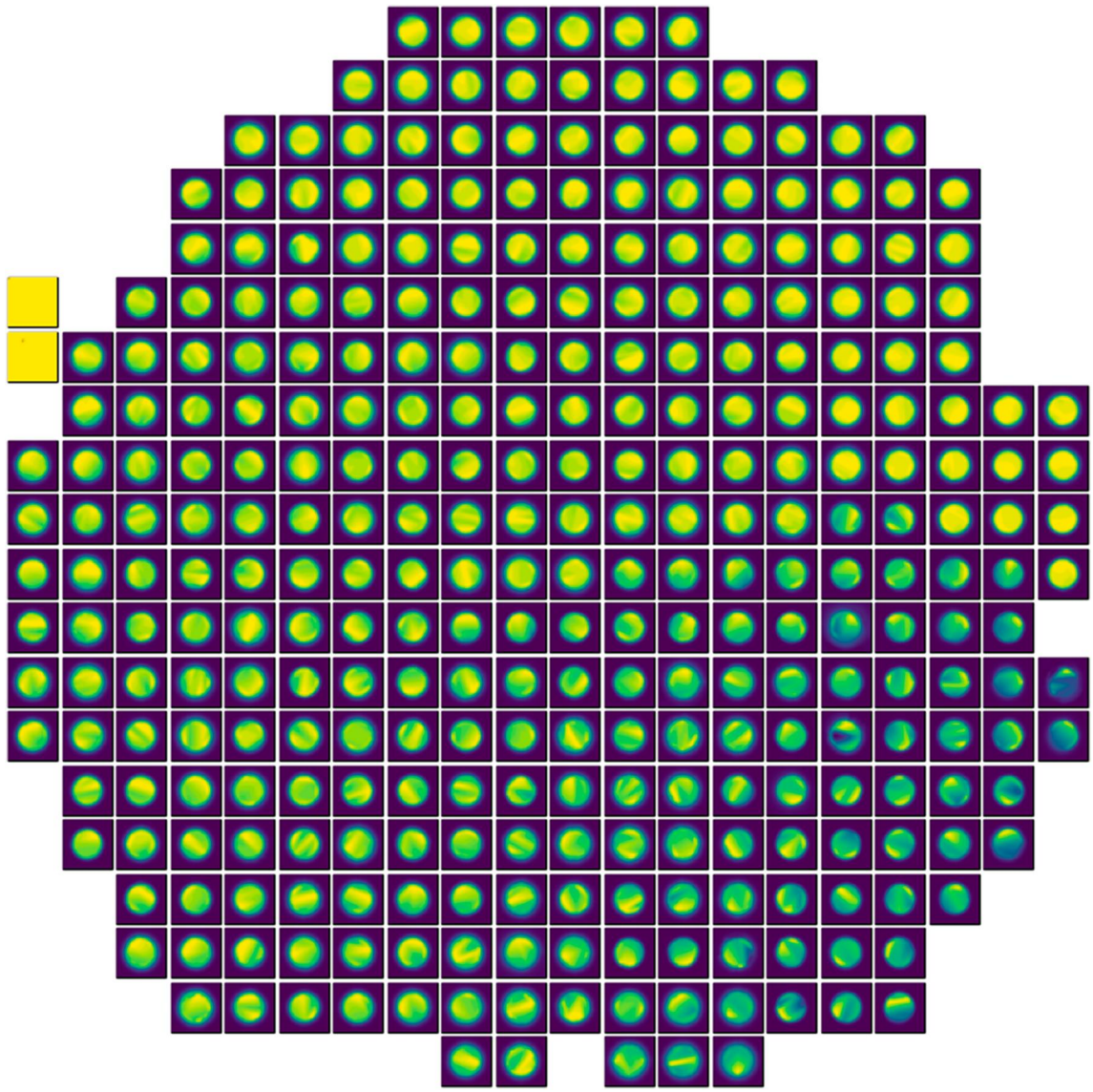


Figure 3. Two dimensional t-SNE derived distributions of image embedding for MMS electron distributions.

distributions for which the processing did not work, several of these images can be seen toward the top left of the visualization (as solid yellow squares) - the method has grouped such anomalous distributions together, providing a method to recognize and remove such distributions. The equivalent representation for the auroral image dataset can be found in the Text S5 in Supporting Information S1.

3.2. Similarity Search

Whilst qualitatively useful to check the embedding process, Figure 3 only shows a reduced dimensional representation of the embeddings. More powerfully, we can use the relative distance in the 512 dimensional embedding space to evaluate the similarity of images. We use a Euclidean distance metric to perform this comparison, we will discuss this choice in Section 4. We note that both datasets number nearly 6,000 examples and have been shuffled prior to generating the embeddings.

Figure 4 shows the six nearest neighbors (A–F) of a "seed" image for the electron distributions. This seed has been taken from an electron diffusion region on 10th August 2017 that has been previously studied (Zhou et al., 2019; Hasegawa et al., 2022; Li et al., 2019). By eye, the images A–F clearly show similarity to this distribution.

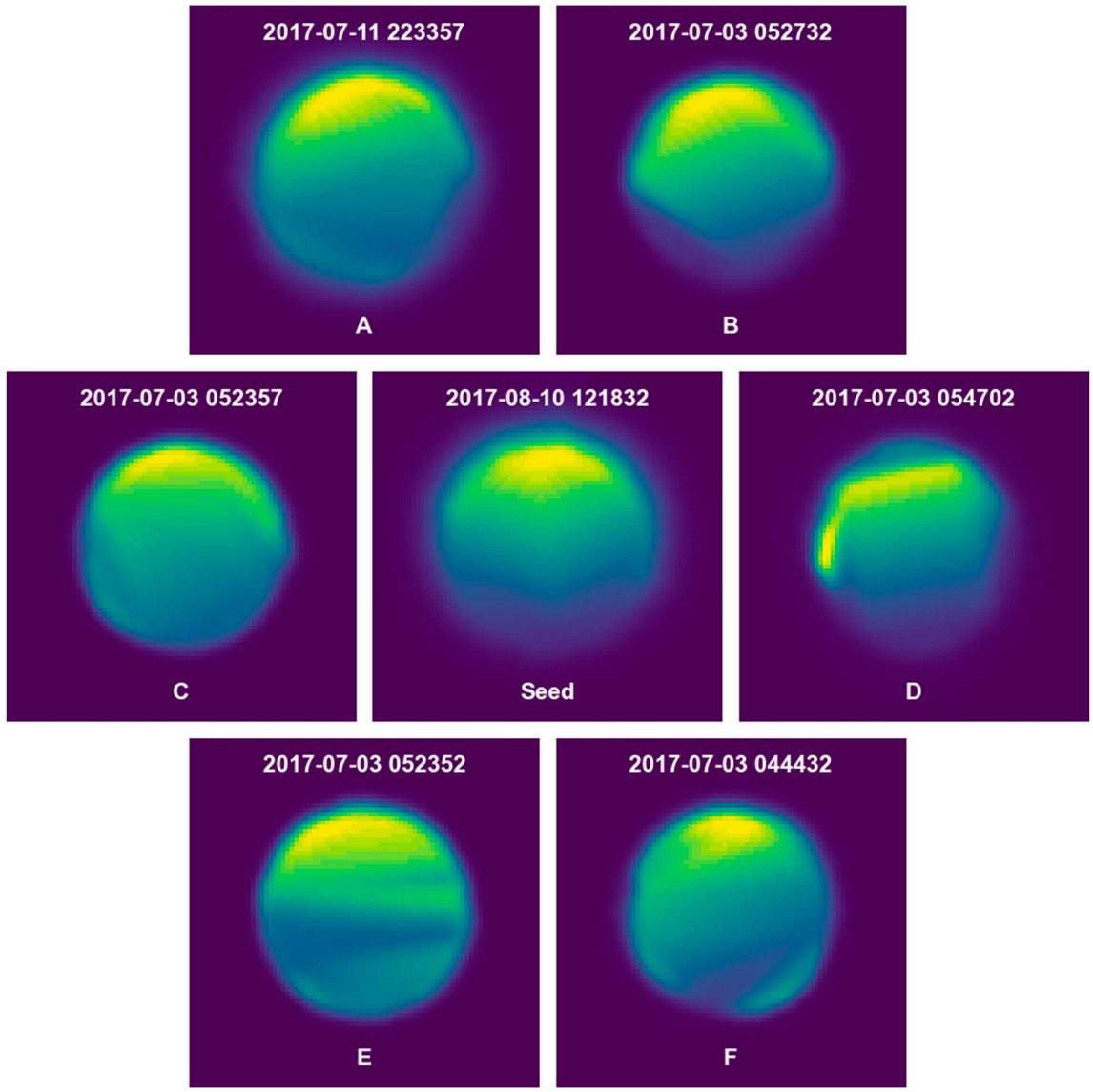


Figure 4. The results of searching for the (Euclidean) nearest neighbors of a "seed" image in the 512 dimensional embedding space for MMS electron distributions. The initial seed image is provided at the center, while the six nearest neighbors are arranged in a circle (A–F) surrounding the seed.

Importantly, we note that the two closest images (A and B) correspond to other electron diffusion regions identified by previous studies: A is close to the one identified by Torbert et al. (2018), while B is proximate to one identified by Chen et al. (2019). This demonstrates the utility of this method in providing similar examples that may show the phenomena of interest, and that this method works well with rare images. In this case, if we had only identified the rare "seed" observation, this method would have enabled us to quickly identify other instances of this phenomenon.

Meanwhile, for the auroral all sky images, in Figure 5 we provided an example auroral arc with interesting small scale wave-like structure. Four of the closest images to this example (A, B, D and E) are from the same observatory (SNKQ) and are within several minutes of the example. However, image C is from an adjacent station (GILL, to the west of SNKQ) but of the same arc, within 90 s of the original image. Our method has clearly found the key, distinctive features of this arc. By eye these two images are not identical—yet the method has found distinct structure that is common to both auroral arcs. Interestingly, image F is from a different station 8 years later showing an arc with some interesting, similar structure.

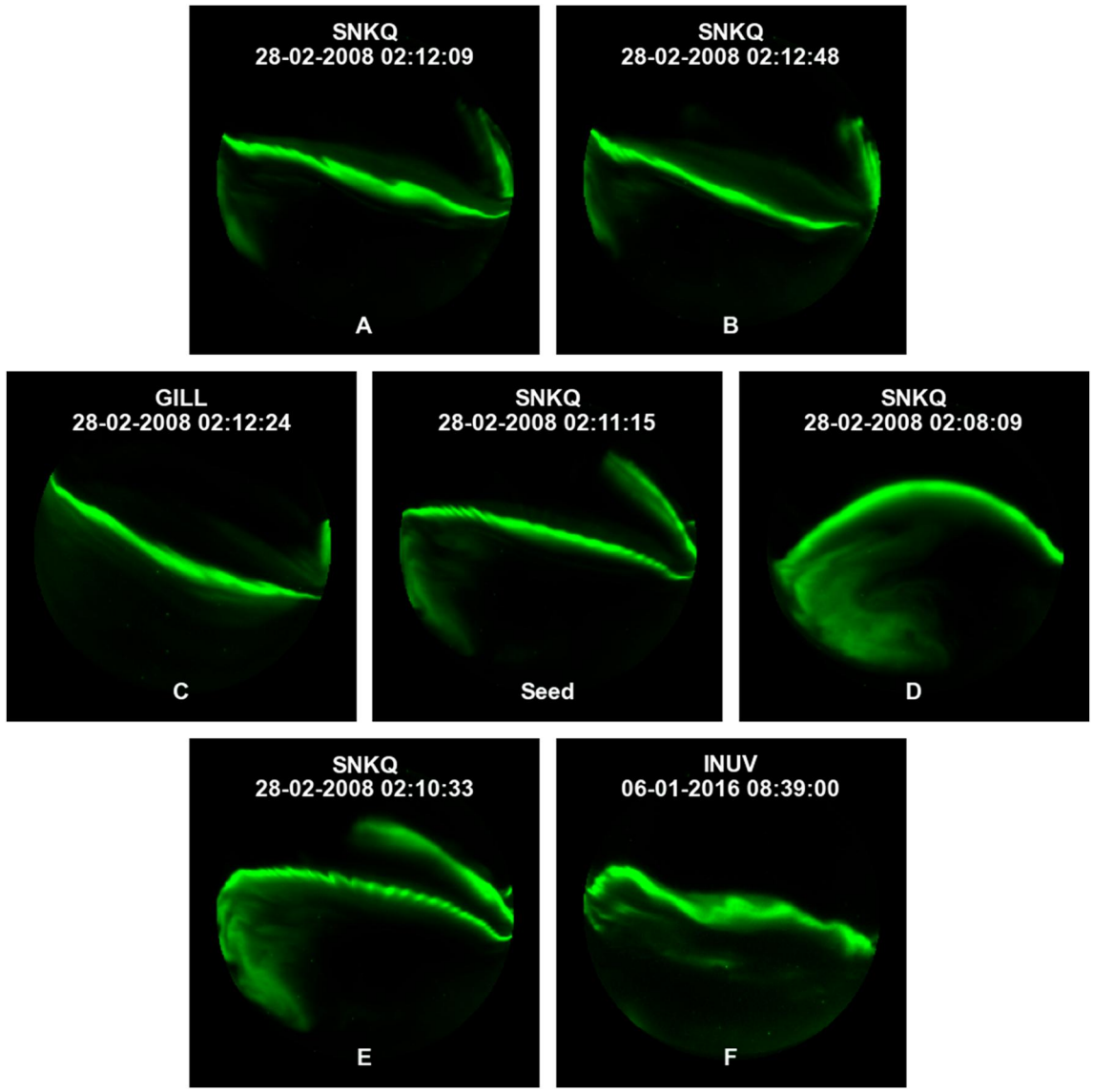


Figure 5. The results of searching for the (Euclidean) nearest neighbors of a "seed" image in the 512 dimensional embedding space for the auroral all sky images. The format is similar to Figure 4.

4. Discussion

4.1. Designing Embeddings

There are two critical steps that determine the embeddings that will be produced by the method: (a) the data processing steps, and (b) the choice of image augmentations. First, the data must be processed into images. These images must be generated such that the feature of interest is clear, for example, the processing to highlight the crescent features in the MMS electron distribution data (Text S1 in Supporting Information S1). Equally, whilst in this work we have used lightly processed all sky images (e.g., Clausen & Nickisch, 2018), mapping such images to a fixed altitude (e.g., 110 km) would likely allow the method to focus more clearly on auroral structure, mitigating the impact of the fish-eye lens.

Second, the selected image augmentations will ultimately determine what features within the image the embedding will ignore. For example, including random rotations would result in the method being insensitive to the relative orientation of features in the image, while blurring will reduce the embeddings dependence upon small

scale structure. These choices can be made relatively simply by considering both the data and the research purpose for which the method is being used.

However, while the necessary augmentations and data processing can be straightforwardly selected, we note that the embedding may include features of the original image beyond the original intention, and which may not match with those that would be selected by eye. For example, in Figure 5 we can see that the closest matches to the seed image also include auroral forms around the edge of the field of view, for example, wrapping around the right and left of Figures 5A and 5B. If the focus is on the small scale features of the aurora at zenith, then additional preprocessing (e.g., cropping) would need to be included.

4.2. Relative Embedding Distances

In this work we explicitly assume observations are similar if the Euclidean distance between their 512 dimension embedding and the target is small. However, we emphasize that these dimensions do not have a physical counterpart, and this is an assumption that we make a priori. The dimensions and ultimately the Euclidean distances will depend on the steps described above: the image processing and the selected augmentations. Additionally, while we have treated the "collapse" of the model (where it generates the same or similar output vectors from every image) as a single, undesirable outcome, collapse can occur on a spectrum (Li et al., 2022), which will impact the Euclidean distances between embeddings, and the utilization of the embedding space.

Nonetheless, the choice of distance metric and assumptions of similarity in a reduced dimensional space is a problem that impacts many recent unsupervised techniques (e.g., Bakrania et al., 2020). This is perhaps in contrast to techniques such as that of Stephens et al. (2019, 2023), who align similar observations based upon a "magnetospheric state vector" derived from physical solar wind parameters and magnetospheric indices, where it is possible to physically interpret the distance.

4.3. Future Utility

The method employed is simple, ultimately only requiring a folder of images from which to train. The choices made prior to the training process mostly amount to image processing and cropping, and selecting appropriate augmentations for the data and task.

The most straightforward future use of the technique is to autonomously search for examples that are similar to a provided case study. This would quickly provide other events that can then be used to compare and contrast with an original, manually identified example. We have directly shown this capability in Figure 4, where we recover other known examples of rare events. In principle this can be performed on any two dimensional data, for example, to search for reconnection separatrices (in a different projection of the electron velocity distributions: Bai et al. (2019)). Nonetheless, care must be taken as the embeddings may be dependent upon irrelevant parts of the image, such as noise or features at the edge of the field of view (e.g., as highlighted in Figure 5).

The pre-trained model could also be fine-tuned with a small labeled dataset for classification tasks, reducing the time required to generate labeled data. The use of self-supervised models in this way has shown excellent performance compared to supervised techniques, in both general tasks (Chen & He, 2021) as well as those related to space plasma physics (Johnson et al., 2021).

The embeddings could also be clustered, in a similar fashion to how unsupervised clustering has been used in magnetospheric physics, for example, to identify the sources of solar wind plasma (Amaya et al., 2020; Bloch et al., 2020), or spatial regions of the magnetosphere (e.g., Bakrania et al., 2020; Innocenti et al., 2021). We note that the large dimensionality of the embedding (512 in our case), indicates it likely contains extra information beyond that required in such a use case. If so, the dimensionality could be further reduced through other methods (e.g., Principle Component Analysis or Autoencoders).

The embeddings could also be as an information-dense input to a more complex model along with other time series data, for example, with the data provided by instruments such as magnetometers. For example, methods aiming to identify electron diffusion regions could use the embedding as an input, in addition to empirically derived scalar quantities based on the electron distributions (e.g., Lenouvel et al., 2021). This would provide the key information contained within 2-dimensional dataset, via the form of bespoke embedding. As discussed above,

our 512 dimensional input may be large compared with other input parameters: care would need to be taken in the design of the model.

5. Conclusions

We have presented two applications of a powerful and easy to use framework by which lower dimensional embeddings can be extracted from large databases of two dimensional data (images). These embeddings are derived through novel self-supervised deep learning models that can extract complex features from two dimensional data, and do not require the time consuming manual labeling of the data.

We have demonstrated the utility of the method through its application to two example datasets: (a) electron distributions, and (b) auroral all sky images. In both cases the method is able to generate embeddings that group similar data. These embeddings can be exploited to reveal the "nearest-neighbors" of a provided sample. In the case of the electron distributions, when provided with an example of a rare type of distribution the technique correctly returns two other instances of this same rare phenomenon from a catalog of nearly 6,000 images. Meanwhile, when provided with an example auroral structure the method locates images taken within a few minutes of the original (given the high auto-correlation of the images), as well as a simultaneous observation of the same auroral structure from a different location.

This technique represents a paradigm shift. It is broadly applicable, and promises to reduce the time taken to manually label data, allowing similar events to be automatically identified and enable the use of two dimensional data as an input to deep learning classification methods.

Data Availability Statement

The MMS data presented in this work are the L2 data and can be accessed from MMS Science Data Center (https://lasp.colorado.edu/mms/sdc/public/), or using SPEDAS (Angelopoulos et al., 2019). The all sky image data are available from the THEMIS website (http://themis.ssl.berkeley.edu), but the data used here are from Clausen and Nickisch (2018). The analysis in this paper was performed using python, including the pandas (McKinney, 2010), NumPy (Van Der Walt et al., 2011), SciPy (Virtanen et al., 2020), Matplotlib (Hunter, 2007), PyTorch (Paszke et al., 2019) and Lightly (Susmelj et al., 2020) libraries. Code to train a SimSIAM model is provided in Smith (2023), and also here: https://github.com/SmithAndy005/SpaceSSL.

Acknowledgments
We would like to thank SpaceML for the inspiration behind the work (Koul et al., 2020). Further, we thank Clausen and Nickisch (2018) for their diligent processing, labeling and provision of THEMIS all sky image data. We would also like to thank Lightly (Susmelj et al., 2020) for their informative tutorials (https://docs.lightly.ai/self-supervised-learning/index.html). AWS was supported by NERC Independent Research Fellowship NE/W009129/1. JES was supported by the Royal Society University Research Fellowship URF/R1/201286. WJS was supported by NASA Grant 80NSSC21K0517.

References

Agastya, C., Ghebremusse, S., Anderson, I., Reed, C., Vahabi, H., & Todeschini, A. (2021). Self-supervised contrastive learning for irrigation detection in satellite imagery. In *Tackling Climate Change with Machine Learning Workshop at ICML 2021*. Retrieved from http://arxiv.org/abs/2108.05484

Akasofu, S.-I. (1964). The development of the auroral substorm. *Planetary and Space Science*, *12*(4), 273–282. https://doi.org/10.1016/0032-0633(64)90151-5

Alosaimi, N., Alhichri, H., Bazi, Y., Ben Youssef, B., & Alajlan, N. (2023). Self-supervised learning for remote sensing scene classification under the few shot scenario. *Scientific Reports*, *13*(1), 1–15. https://doi.org/10.1038/s41598-022-27313-5

Amaya, J., Dupuis, R., Innocenti, M. E., & Lapenta, G. (2020). Visualizing and interpreting unsupervised solar wind classifications. *Frontiers in Astronomy and Space Sciences*, *7*, 66. https://doi.org/10.3389/FSPAS.2020.553207/BIBTEX

Angelopoulos, V., Cruce, P., Drozdov, A., Grimes, E. W., Hatzigeorgiu, N., King, D. A., et al. (2019). The Space Physics Environment Data Analysis System (SPEDAS) [Software]. Space Science Reviews, *215*(1), 9. https://doi.org/10.1007/s11214-018-0576-4

Bai, S. C., Shi, Q., Zong, Q. G., Wang, X., Tian, A., Degeling, A. W., et al. (2019). Electron dispersion and parallel electron beam observed near the separatrix. *Journal of Geophysical Research: Space Physics*, *124*(9), 7494–7504. https://doi.org/10.1029/2019JA026836

Bakrania, M. R., Rae, I. J., Walsh, A. P., Verscharen, D., & Smith, A. W. (2020). Using dimensionality reduction and clustering techniques to classify space plasma regimes. *Frontiers in Astronomy and Space Sciences*, *7*, 80. https://doi.org/10.3389/fspas.2020.593516

Bergstedt, K., Ji, H., Jara-Almonte, J., Yoo, J., Ergun, R. E., & Chen, L. J. (2020). Statistical properties of magnetic structures and energy dissipation during turbulent reconnection in the Earth's magnetotail. *Geophysical Research Letters*, *47*(19), e2020GL088540. https://doi.org/10.1029/2020GL088540

Bloch, T., Watt, C., Owens, M., McInnes, L., & Macneil, A. R. (2020). Data-driven classification of coronal hole and streamer belt solar wind. *Solar Physics*, *295*(3), 41. https://doi.org/10.1007/s11207-020-01609-z

Burch, J. L., Moore, T. E., Torbert, R. B., & Giles, B. L. (2016). Magnetospheric multiscale overview and science objectives. *Space Science Reviews*, *199*(1–4), 5–21. https://doi.org/10.1007/s11214-015-0164-9

Burch, J. L., Torbert, R. B., Phan, T. D., Chen, L.-J., Moore, T. E., Ergun, R. E., et al. (2016). Electron-scale measurements of magnetic reconnection in space. *Science*, *352*(6290), aaf2939. https://doi.org/10.1126/science.aaf2939

Chen, L. J., Hesse, M., Wang, S., Gershman, D., Ergun, R. E., Burch, J., et al. (2017). Electron diffusion region during magnetopause reconnection with an intermediate guide field: Magnetospheric multiscale observations. *Journal of Geophysical Research: Space Physics*, *122*(5), 5235–5246. https://doi.org/10.1002/2017JA024004

Chen, L. J., Wang, S., Hesse, M., Ergun, R. E., Moore, T., Giles, B., et al. (2019). Electron diffusion regions in magnetotail reconnection under varying guide fields. *Geophysical Research Letters*, *46*(12), 6230–6238. https://doi.org/10.1029/2019GL082393

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *37th International Conference on Machine Learning, ICML 2020* (Vol. *PartF16814*, pp. 1575–1585). Retrieved from https://github.com/google-research/simclr

Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. Retrieved from https://arxiv.org/abs/2003.04297

Chen, X., & He, K. (2021). Exploring simple Siamese representation learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 15745–15753). https://doi.org/10.1109/CVPR46437.2021.01549

Clausen, L. B., & Nickisch, H. (2018). Automatic classification of auroral images from the Oslo Auroral THEMIS (OATH) data set using machine learning [Dataset]. Journal of Geophysical Research: Space Physics, *123*(7), 5640–5647. https://doi.org/10.1029/2018JA025274

Donovan, E., Mende, S., Jackel, B., Frey, H., Syrjäsuo, M., Voronkov, I., et al. (2006). The THEMIS all-sky imaging array—System design and initial results from the prototype imager. *Journal of Atmospheric and Solar-Terrestrial Physics*, *68*(13), 1472–1487. https://doi.org/10.1016/J.JASTP.2005.03.027

Eastwood, J. P., Phan, T. D., Øieroset, M., & Shay, M. A. (2010). Average properties of the magnetic reconnection ion diffusion region in the Earth's magnetotail: The 2001-2005 Cluster observations and comparison with simulations. *Journal of Geophysical Research: Space Physics*, *115*(8). https://doi.org/10.1029/2009JA014962

Endo, T., & Matsumoto, M. (2022). Aurora image classification with deep metric learning. *Sensors*, *22*(17), 6666. https://doi.org/10.3390/S22176666

Farrugia, C. J., Rogers, A. J., Torbert, R. B., Genestreti, K. J., Nakamura, T. K., Lavraud, B., et al. (2021). An encounter with the ion and electron diffusion regions at a flapping and twisted tail current sheet. *Journal of Geophysical Research: Space Physics*, *126*(3), e2020JA028903. https://doi.org/10.1029/2020JA028903

Forsyth, C., Sergeev, V. A., Henderson, M. G., Nishimura, Y., & Gallardo-Lacourt, B. (2020). Physical processes of meso-scale, dynamic auroral forms. *Space Science Reviews*, *216*(4), 46. https://doi.org/10.1007/s11214-020-00665-y

Gao, C. H., Tang, B. B., Guo, X. C., Li, W. Y., Khotyaintsev, Y. V., Graham, D. B., et al. (2023). Agyrotropic electron distributions in the terrestrial foreshock transients. *Geophysical Research Letters*, *50*(4), e2022GL102235. https://doi.org/10.1029/2022GL102235

Garton, T. M., Jackman, C. M., Smith, A. W., Yeakel, K. L., Maloney, S. A., & Vandegriff, J. (2021). Machine learning applications to Kronian magnetospheric reconnection classification. *Frontiers in Astronomy and Space Sciences*, *7*, 104. https://doi.org/10.3389/fspas.2020.600031

Graham, D. B., Khotyaintsev, Y. V., Vaivads, A., Norgren, C., André, M., Webster, J. M., et al. (2017). Instability of agyrotropic electron beams near the electron diffusion region. *Physical Review Letters*, *119*(2), 025101. https://doi.org/10.1103/PHYSREVLETT.119.025101/FIGURES/4/MEDIUM

Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., et al. (2020). Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems, 2020-Decem*. Retrieved from https://arxiv.org/abs/2006.07733v3

Hamrin, M., Andersson, L., Vaivads, A., Pitkänen, T., & Gunell, H. (2015). The use of the power density for identifying reconnection regions. *Journal of Geophysical Research A: Space Physics*, *120*(10), 8644–8662. https://doi.org/10.1002/2015JA021535

Hasegawa, H., Denton, R. E., Nakamura, T. K. M., Genestreti, K. J., Phan, T. D., Nakamura, R., et al. (2022). Magnetic field annihilation in a magnetotail electron diffusion region with electron-scale magnetic island citation. *Journal of Geophysical Research: Space Physics*, *127*(7). https://doi.org/10.1029/2022JA030408

Hayat, M. A., Stein, G., Harrington, P., Lukić, Z., & Mustafa, M. (2021). Self-supervised representation learning for astronomical images. *The Astrophysical Journal Letters*, *911*(2), L33. https://doi.org/10.3847/2041-8213/ABF2C7

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2019). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 9726–9735). https://doi.org/10.1109/CVPR42600.2020.00975

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem* (pp. 770–778). https://doi.org/10.48550/arxiv.1512.03385

Henderson, M. G. (2013). Auroral substorms, poleward boundary activations, auroral streamers, omega bands, and onset precursor activity. In *Auroral Phenomenology and Magnetospheric Processes: Earth and Other Planets* (pp. 39–54). https://doi.org/10.1029/2011GM001165

Hinton, G., & Roweis, S. (2003). Stochastic neighbor embedding. In *Advances in Neural Information Processing Systems*. https://doi.org/10.1007/978-3-031-10602-6_16

Huang, S., Jiang, K., Zhou, M., Huang, S. Y., Jiang, K., Yuan, Z. G., et al. (2018). Observations of the electron jet generated by secondary reconnection in the terrestrial magnetotail reconnection view project MMS magnetotail reconnection event of July 11 2017 view project observations of the electron jet generated by secondary reconnect. *The Astrophysical Journal*, *862*(2), 144. https://doi.org/10.3847/1538-4357/aacd4c

Huang, S., Zhao, P., He, J., Yuan, Z., Zhou, M., Fu, H., et al. (2018). A new method to identify flux ropes in space plasmas. *Annales Geophysicae*, *36*(5), 1275–1283. https://doi.org/10.5194/ANGEO-36-1275-2018

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment [Software]. Computing in Science & Engineering, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Innocenti, M. E., Amaya, J., Raeder, J., Dupuis, R., Ferdousi, B., & Lapenta, G. (2021). Unsupervised classification of simulated magnetospheric regions. *Annales Geophysicae*, *39*(5), 861–881. https://doi.org/10.5194/angeo-39-861-2021

Johnson, J. W., Hari, S., Hampton, D., Connor, H. K., & Keesee, A. (2021). A contrastive learning approach to auroral identification and classification. In *Proceedings - 20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021* (pp. 772–777). https://doi.org/10.1109/ICMLA52953.2021.00128

Kalmoni, N. M. E., Rae, I. J., Murphy, K. R., Forsyth, C., Watt, C. E., & Owen, C. J. (2017). Statistical azimuthal structuring of the substorm onset arc: Implications for the onset mechanism. *Geophysical Research Letters*, *44*(5), 2078–2087. https://doi.org/10.1002/2016GL071826

Killey, S., Rae, I. J., Chakraborty, S., Smith, A. W., Bentley, S. N., Bakrania, M. R., et al. (2023). Using machine learning to diagnose relativistic electron distributions in the Van Allen radiation belts. *RAS Techniques and Instruments*, *2*(1), 548–561. https://doi.org/10.1093/rasti/rzad035

Koul, A., Ganju, S., Kasam, M., & Parr, J. (2020). SpaceML: Distributed open-source research with citizen scientists for the advancement of space technology for NASA. Retrieved from https://arxiv.org/abs/2012.10610v3

Lenouvel, Q., Génot, V., Garnier, P., Toledo-Redondo, S., Lavraud, B., Aunai, N., et al. (2021). Identification of electron diffusion regions with a machine learning approach on MMS data at the Earth's magnetopause. *Earth and Space Science*, *8*(5), e2020EA001530. https://doi.org/10.1029/2020EA001530

Li, A. C., Efros, A. A., & Pathak, D. (2022). Understanding collapse in non-contrastive siamese representation learning. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13691 LNCS* (pp. 490–505). https://doi.org/10.1007/978-3-031-19821-2_28/COVER

Li, X., Wang, R., & Lu, Q. (2023). Division of magnetic flux rope via magnetic reconnection observed in the magnetotail. *Geophysical Research Letters*, *50*(1), e2022GL101084. https://doi.org/10.1029/2022GL101084

Li, X., Wang, R., Lu, Q., Hwang, K. J., Zong, Q., Russell, C. T., & Wang, S. (2019). Observation of nongyrotropic electron distribution across the electron diffusion region in the magnetotail reconnection. *Geophysical Research Letters*, *46*(24), 14263–14273. https://doi.org/10.1029/2019GL085014

Liu, H., HaoChen, J. Z., Gaidon, A., & Ma, T. (2022). SELF-SUPERVISED learning IS more robust to dataset imbalance. In *ICLR 2022 - 10th International Conference on Learning Representations*

Louis, C., Jackman, C., Mangham, S., Smith, K., O'Dwyer, E., Empey, A., et al. (2022). The "SPectrogram Analysis and Cataloguing Environment" (SPACE) labelling tool. *Frontiers in Astronomy and Space Sciences*, *9*, 332. https://doi.org/10.3389/FSPAS.2022.1001166/BIBTEX

McKinney, W. (2010). Data structures for statistical computing in Python [Software]. Retrieved from http://conference.scipy.org/proceedings/scipy2010/mckinney.html

Nishimura, Y., Lyons, L. R., Angelopoulos, V., Kikuchi, T., Zou, S., & Mende, S. B. (2011). Relations between multiple auroral streamers, pre-onset thin arc formation, and substorm auroral onset. *Journal of Geophysical Research: Space Physics*, *116*(A9), 9214. https://doi.org/10.1029/2011JA016768

Ohtani, S., & Motoba, T. (2022). Formation of beading auroral arcs at substorm onset: Implications of its variability into the generation process. *Journal of Geophysical Research: Space Physics*, *128*(1), e2022JA030796. https://doi.org/10.1029/2022JA030796

Øieroset, M., Phan, T. D., Fujimoto, M., Lin, R. P., & Lepping, R. P. (2001). In situ detection of collisionless reconnection in the Earth's magnetotail. *Nature*, *412*(6845), 414–417. https://doi.org/10.1038/35086520

Opgenoorth, H. J., Oksman, J., Kaila, K. U., Nielsen, E., & Baumjohann, W. (1983). Characteristics of eastward drifting omega bands in the morning sector of the auroral oval. *Journal of Geophysical Research: Space Physics*, *88*(A11), 9171–9185. https://doi.org/10.1029/JA088IA11P09171

Partamies, N., Weygand, J. M., & Juusola, L. (2017). Statistical study of auroral omega bands. *Annales Geophysicae*, *35*(5), 1069–1083. https://doi.org/10.5194/angeo-35-1069-2017

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury Google, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library [Software]. Advances in Neural Information Processing Systems, *32*. https://doi.org/10.48550/arXiv.1912.01703

Phan, T. D., Eastwood, J. P., Shay, M. A., Drake, J. F., Sonnerup, B. U. Ö., Fujimoto, M., et al. (2018). Electron magnetic reconnection without ion coupling in Earth's turbulent magnetosheath. *Nature*, *557*(7704), 202–206. https://doi.org/10.1038/s41586-018-0091-5

Pollock, C., Moore, T., Jacques, A., Burch, J., Gliese, U., Saito, Y., et al. (2016). Fast plasma investigation for magnetospheric multiscale. *Space Science Reviews*, *199*(1–4), 331–406. https://doi.org/10.1007/S11214-016-0245-4/TABLES/7

Rogers, A. J., Farrugia, C. J., & Torbert, R. B. (2019). Numerical algorithm for detecting ion diffusion regions in the geomagnetic tail with applications to MMS tail season 1 May to 30 September 2017. *Journal of Geophysical Research: Space Physics*, *124*(8), 6487–6503. https://doi.org/10.1029/2018JA026429

Rogers, A. J., Farrugia, C. J., Torbert, R. B., & Rogers, T. J. (2023). Applying magnetic curvature to MMS data to identify thin current sheets relative to tail reconnection. *Journal of Geophysical Research: Space Physics*, *128*(1), e2022JA030577. https://doi.org/10.1029/2022JA030577

Sarkango, Y., Slavin, J. A., Jia, X., DiBraccio, G. A., Clark, G. B., Sun, W., et al. (2022). Properties of ion-inertial scale plasmoids observed by the Juno spacecraft in the Jovian magnetotail. *Journal of Geophysical Research: Space Physics*, *127*(3), e2021JA030181. https://doi.org/10.1029/2021JA030181

Sergeev, V. A., Liou, K., Meng, C. I., Newell, P. T., Brittnacher, M., Parks, G., & Reeves, G. D. (1999). Development of auroral streamers in association with localized impulsive injections to the inner magnetotail. *Geophysical Research Letters*, *26*(3), 417–420. https://doi.org/10.1029/1998GL900311

Sergeev, V. A., Liou, K., Newell, P. T., Ohtani, S. I., Hairston, M. R., & Rich, F. (2004). Auroral streamers: Characteristics of associated precipitation, convection and field-aligned currents. *Annales Geophysicae*, *22*(2), 537–548. https://doi.org/10.5194/angeo-22-537-2004

Shang, Z., Yao, Z., Liu, J., Xu, L., Xu, Y., Zhang, B., et al. (2023). Automated classification of auroral images with deep neural networks. *Universe*, *9*(2), 96. https://doi.org/10.3390/UNIVERSE9020096

Smith, A. W. (2023). *SmithAndy005/SpaceSSL: First Release*. Zenodo. https://doi.org/10.5281/ZENODO.10245627

Smith, A. W., Rae, I. J., Forsyth, C., Watt, C. E. J., & Murphy, K. R. (2020). On the magnetospheric ULF wave counterpart of substorm onset. *Journal of Geophysical Research: Space Physics*, *125*(4). https://doi.org/10.1029/2019JA027573

Smith, A. W., Slavin, J. A., Jackman, C. M., Fear, R. C., Poh, G. K., DiBraccio, G. A., et al. (2017). Automated force-free flux rope identification. *Journal of Geophysical Research: Space Physics*, *122*(1), 780–791. https://doi.org/10.1002/2016JA022994

Stephens, G. K., Sitnov, M. I., Korth, H., Tsyganenko, N. A., Ohtani, S., Gkioulidou, M., & Ukhorskiy, A. Y. (2019). Global empirical picture of magnetospheric substorms inferred from multi-mission magnetometer data. *Journal of Geophysical Research: Space Physics*, *124*(2), 1085–1110. https://doi.org/10.1029/2018JA025843

Stephens, G. K., Sitnov, M. I., Weigel, R. S., Turner, D. L., Tsyganenko, N. A., Rogers, A. J., et al. (2023). Global structure of magnetotail reconnection revealed by mining space magnetometer data. *Journal of Geophysical Research: Space Physics*, *128*(2). https://doi.org/10.1029/2022JA031066

Susmelj, I., Heller, M., Wirth, P., Prescott, J., Ebner, M., & Al, E. (2020). Lightly [Software]. GitHub. Retrieved from https://github.com/lightly-ai/lightly

Syrjasuo, M., Donovan, E., Qin, X., & Yang, Y.-H. (2007). Automatic classification of auroral images in substorm studies. *International Conference on Substorms*(Vol. *1*), 309–313. Retrieved from http://aurora.phys.ucalgary.ca/cbir/. http://www.ava.fmi.fi/~syrjasuo/Data/syrjaesuo_ics8.pdf

Syrjäsuo, M. T., & Donovan, E. F. (2002). Analysis of auroral images: Detection and tracking. *Geophysica*, *38*(1–2), 3–14.

Syrjäsuo, M. T., & Donovan, E. F. (2004). Diurnal auroral occurrence statistics obtained via machine vision. *Annales Geophysicae*, *22*(4), 1103–1113. https://doi.org/10.5194/ANGEO-22-1103-2004

Tian, S., Lyons, L. R., Nishimura, Y., Wygant, J. R., Lysak, R. L., Ferradas, C. P., et al. (2022). Auroral beads in conjunction with kinetic Alfvén waves in the equatorial inner-magnetosphere. *Geophysical Research Letters*, *49*(9), e2022GL098457. https://doi.org/10.1029/2022GL098457

Torbert, R. B., Burch, J. L., Argall, M. R., Alm, L., Farrugia, C. J., Forbes, T. G., et al. (2017). Structure and dissipation characteristics of an electron diffusion region observed by MMS during a rapid, normal-incidence magnetopause crossing. *Journal of Geophysical Research: Space Physics*, *122*(12), 11901–11916. https://doi.org/10.1002/2017JA024579

Torbert, R. B., Burch, J. L., Giles, B. L., Gershman, D., Pollock, C. J., Dorelli, J., et al. (2016). Estimates of terms in Ohm's law during an encounter with an electron diffusion region. *Geophysical Research Letters*, *43*(12), 5918–5925. https://doi.org/10.1002/2016GL069553

Torbert, R. B., Burch, J. L., Phan, T. D., Hesse, M., Argall, M. R., Shuster, J., et al. (2018). Electron-scale dynamics of the diffusion region during symmetric magnetic reconnection in space. *Science*, *362*(6421), 1391–1395. https://doi.org/10.1126/science.aat2998

Van Der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2625.

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation [Software]. Computing in Science & Engineering, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python [Software]. Nature Methods, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wild, J. A., Yeoman, T. K., Eglitis, P., & Opgenoorth, H. J. (2000). Multi-instrument observations of the electric and magnetic field structure of omega bands. *Annales Geophysicae*, *18*(1), 99–110. https://doi.org/10.1007/S00585-000-0099-6

Zhong, Y., Ye, R., Liu, T., Hu, Z., & Zhang, L. (2020). Automatic aurora image classification framework based on deep learning for occurrence distribution analysis: A case study of all-sky image data sets from the Yellow Rriver Station. *Journal of Geophysical Research: Space Physics*, *125*(9), e2019JA027590. https://doi.org/10.1029/2019JA027590

Zhong, Z. H., Zhou, M., Tang, R. X., Deng, X. H., Khotyaintsev, Y. V., Giles, B. L., et al. (2020). Extension of the electron diffusion region in a guide field magnetic reconnection at magnetopause. *The Astrophysical Journal Letters*, *892*(1), L5. https://doi.org/10.3847/2041-8213/AB7B7C

Zhou, M., Deng, X. H., Zhong, Z. H., Pang, Y., Tang, R. X., El-Alaoui, M., et al. (2019). Observations of an electron diffusion region in symmetric reconnection with weak guide field. *The Astrophysical Journal*, *870*(1), 34. https://doi.org/10.3847/1538-4357/AAF16F