
Correlation of Auroral Dynamics and GNSS Scintillation with an Autoencoder

Kara D. Lamb*
CIRES/NOAA
Boulder, CO, USA

Garima Malhotra*
University of Michigan
Ann Arbor, Mi, USA

Athanasios Vlontzos*
Imperial College London
London, UK

Edward Wagstaff*
University of Oxford
Oxford, UK

Atılım Günes Baydin
University of Oxford
Oxford, UK

Anahita Bhiwandiwalla
Intel AI Lab
Santa Clara, CA, USA

Yarin Gal
University of Oxford
Oxford, UK

Alfredo Kalaitzis
Element AI
London, UK

Anthony Reina
Intel AIPG
San Diego, CA, USA

Asti Bhatt
SRI International
Menlo Park, CA, USA

Abstract

High energy particles originating from solar activity travel along the Earth's magnetic field and interact with the atmosphere around the higher latitudes. These interactions often manifest as aurora in the form of visible light in the Earth's ionosphere. These interactions also result in irregularities in the electron density, which cause disruptions in the amplitude and phase of the radio signals from the Global Navigation Satellite Systems (GNSS), known as 'scintillation'. In this paper we use a multi-scale residual autoencoder (Res-AE) to show the correlation between specific dynamic structures of the aurora and the magnitude of the GNSS phase scintillations (σ_ϕ). Auroral images are encoded in a lower dimensional feature space using the Res-AE, which in turn are clustered with t-SNE and UMAP. Both methods produce similar clusters, and specific clusters demonstrate greater correlations with observed phase scintillations. Our results suggest that specific dynamic structures of auroras are highly correlated with GNSS phase scintillations.

1 Introduction

Space weather includes phenomena associated with high energy particles originating from the sun and their expansion throughout the solar system. As the high energy particles interact and are captured by the Earth's magnetic field they are deposited in high latitudes around the two poles. Their increased concentration creates perturbations in the Earth's ionosphere which affect the propagation of radio signal communications depending on the scale of perturbations and the frequency of the signals.

The global navigation satellite system (GNSS) is a constellation of satellites around the Earth used for accurate navigation, timing and positioning. Each GNSS receiver is connected to four or more satellites receiving radio signals used to calculate accurate location and timing coordinates of the receiver. The frequency of these signals is in GHz and therefore, they interact with small scale irregularities in the ionosphere. This can cause the phase and the amplitude of the radio waves to change, resulting in degraded performance or even complete loss of lock. The changes in amplitude and phase are known as *scintillations* – this paper is focused on *phase* scintillations (σ_ϕ), as these are considered to be more important in the high latitudes [12].

*Equal Contributions

At high latitudes, the interaction of high energy solar wind particles with terrestrial neutral gases in the Earth’s ionosphere leads to visually observed effects such as the Aurora Borealis/Australis. Aurorae are highly dynamic phenomena that have long been thought to correlate with amplitude and phase scintillations due to similar geophysical drivers causing these phenomena. Previous research has demonstrated that the occurrence of significant phase scintillations is correlated with the presence of the aurora [1, 12, 13], and that specific structures in the aurora may be more correlated with high σ_ϕ values (e.g. [5, 10, 13, 6]).

Variations in the visible aurora are manifestations of variations in the geophysical drivers. Auroral images have long been classified in a few distinct classes pointing to their geophysical driving mechanisms. Several recent studies have applied deep learning methods to aurora image classification [2, 14, 15]. These studies have used supervised learning approaches, which can be subjective, as they rely on the judgement of human experts to label auroral events. Human annotation can bias labels to identifying only the most prominent of features in the images, and also it relies on pre-defining specific classes that may exclude some of the salient features associated with physical processes occurring in the ionosphere. In addition to being time-consuming, labelling typically focuses on identifying application-specific features. In this work, we investigate unsupervised approaches to clustering aurora images using a general-purpose approach. Images of the Aurora Borealis originating from the THEMIS network of Northern Canada [9] are passed through a Residual Autoencoder (Res-AE) [3, 4], creating high-dimensional embeddings. The embeddings are then dimensionally-reduced and clustered with t-SNE and UMAP [7, 8]. We investigate the correlations of clusters to both known aurora classes and the magnitude of the phase scintillations measured at co-located GNSS receivers.

2 Proposed Method

Dataset Our dataset contains 35,277 images from THEMIS (Time History of Events and Macroscale Interactions during Substorms) All-Sky Imager (ASI) data from the Fort Simpson (FSIM) site in Northern Canada. The images are part of the THEMIS auroral dataset, taken ± 3 hours from midnight and represent two months of data from January and February 2015. An additional 7700 manually annotated auroral images, classified among 6 classes (*arc, diffuse, discrete, moon, clear, clouds*) depending on the auroral structure [2] from March 2015, are used to evaluate embeddings in the low-dimensional feature space. We note that images are classified in the moon category only if the presence of the moon makes the aurora not visible (per the discretion of the expert annotating the images). We also use measurements from a PolarRxS GNSS receiver (Septentrio 2015, Leuven, Belgium) at Fort Simpson from the Canadian High Arctic Ionospheric Network (CHAIN) to provide measurements of the scintillation indices at the same time as the aurora images.

Residual Autoencoder In this work we use a Residual Autoencoder (Res-AE) as introduced in [4]. Figure 1 illustrates the encoder, and the decoder mirrors the encoder. The overall architecture has a U-Net like structure, where sequential convolutions create multi-scale representations of the input [11]. Each resolution level includes a residual block [3] and a convolutional layer with a stride of 2. The resulting feature map is passed to the next resolution level. Furthermore, a series of convolutional layers map the features from all resolution levels to an encoding of size $L_1 \times L_2 \times 128$, adding them all together to create an intermediate representation L' . A final convolutional layer then maps the intermediate representation to the encoding L of size $L_1 \times L_2 \times L_3$. The size of the representation L is a critical hyperparameter as it defines the expressiveness of the encoder. In our work we found the optimal trade-off between feature space compression and expressiveness to be $32 \times 32 \times 3$. The multi-scale and residual nature of the autoencoder compresses the high-dimensional input images into a feature representation that accounts for multi-scale features and dependencies while filtering out some of the noise. Figure 2 shows an example of the test and reconstructed images using the Res-AE trained on images from the FSIM Themis site.

Unsupervised clustering of embedded latent representations We investigate Barnes-Hut t-distributed Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction of the latent space [7, 8]. Both approaches provide lower dimensional representations of higher dimensional data, such that essential topological structures of the higher dimensional space are preserved in the density of embeddings in the lower dimensional space. In our approach, THEMIS images are mapped to the latent space L using the trained Res-AE and then clustered with t-SNE or UMAP to embed the encoded latent representation into two dimensions for easier visualisation and interpretation. In this case we constrain the local

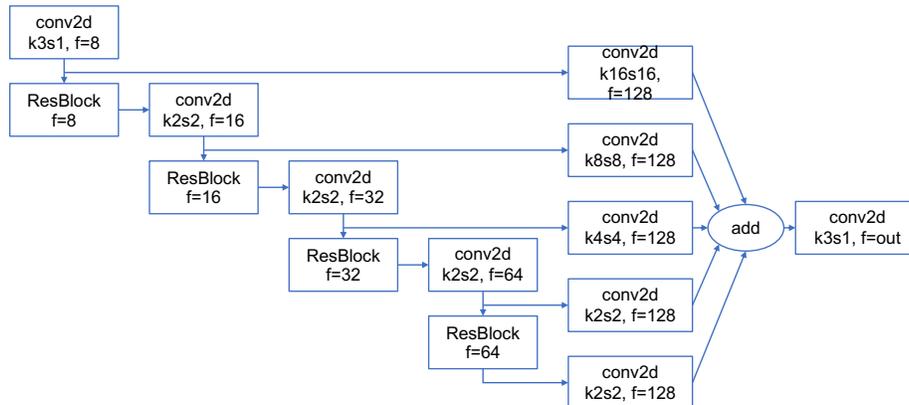


Figure 1: Res-AE Encoder

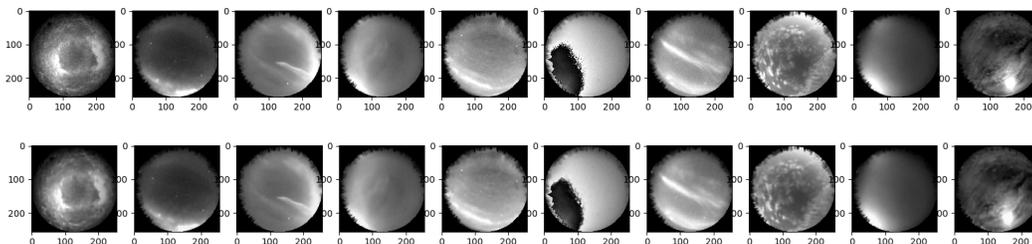


Figure 2: Original (top) and reconstructed (bottom) Themis images.

neighborhood from which UMAP learns the manifold structure of the data to 15 nearest neighbors, as this provides a reasonable separation between the clusters.

3 Results

We train the Res-AE on 2 months of THEMIS images from FSIM to encode our image class labelled dataset of 7700 images (we exclude images labelled as *cloudy*, as these do not contain physical information on the ionosphere). Figure 3 shows the embedded latent representations for both the t-SNE and UMAP projections, colored by the image class labels. Both t-SNE and UMAP generate similar clusters of the labelled image classes from the latent representations. They both most clearly cluster images identified as *moon*, *arc*, and *diffuse*, with subsets of the *discrete* images more closely associated with each of these clusters.

We investigate the relation of the unsupervised image clusters to physically-relevant quantities related to the ionospheric electron density, namely the phase scintillation index (σ_ϕ). Previous work has demonstrated that the intensity in the white light ASI images is correlated with the observed auroral precipitated energy [9], indicating ASI images provide meaningful insight into physical processes occurring in the ionosphere, such as localised fluctuations in electron density contributing to GNSS scintillations. Using spectral clustering, we identify distinct clusters in the projected latent space and look at the log-normal distribution of σ_ϕ measured by GNSS receivers co-located at the same site as the THEMIS ASI imagers. Figure 4a demonstrates this approach on the UMAP projection. Specific clusters of auroral images (e.g. clusters 0 and 5 in Figure 4b) are correlated with significantly higher phase scintillation indices. These clusters correspond to a subset of the aurora classified as

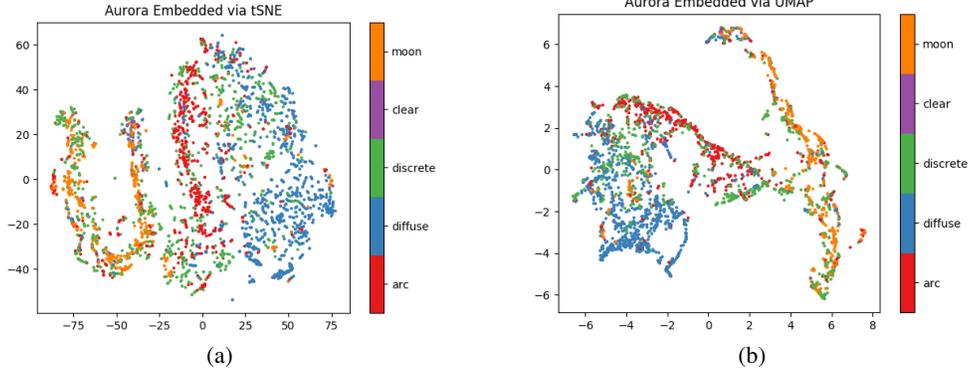


Figure 3: Latent space colored by image classes projected by (a) t-SNE (b) UMAP

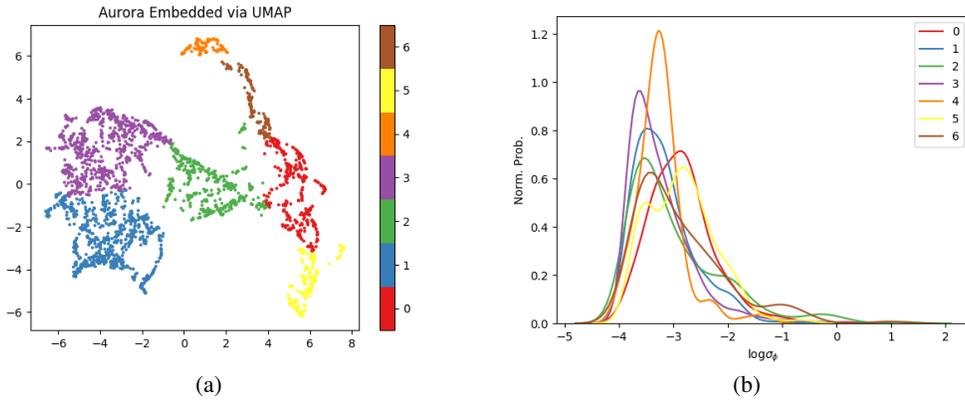


Figure 4: UMAP projection showing (a) spectral clustering (b) the corresponding log-normal distributions of σ_ϕ associated with each cluster. Colors and numbers in (b) correspond to the clusters identified in (a) by the spectral clustering algorithm applied to the UMAP projection.

discrete and *arc* of Figure 3b. These clusters suggest that the brightness of auroral features (and thus the magnitude of the auroral precipitated energy) may be correlated with the magnitude of the σ_ϕ . This analysis demonstrates that clustering in the low dimensional projection of the latent space can provide physically meaningful correlations with relevant physical parameters, such as σ_ϕ .

4 Discussion and Outlook

While previous work has demonstrated that supervised deep learning approaches provide a useful method for analysing THEMIS images, here we demonstrate that unsupervised clustering techniques using a general-purpose Autoencoder may be a useful alternative approach. These results indicate that non-linear dimensionality-reduction techniques such as UMAP and t-SNE can provide meaningful lower-dimensional projections of the latent representation of aurora images that correlate with clusters associated with both human annotated image classes and physically meaningful parameters related to the ionospheric electron density variations. These results also indicate that specific dynamic structures in the aurora (as observed by the ASI) are more likely to correlate with GNSS phase scintillations. Such an approach seems to be site specific however, as we observed greater separation between images measured at different sites versus separation between image classes at a single site; additional masking of image edges may extend the applicability of the method to multiple sites. Future work will focus on applying this method to a larger image data set, further exploration of unsupervised auroral feature extraction from the t-SNE and UMAP projections correlated with high σ_ϕ values, and investigation of the correlation of clusters with other ionospheric parameters, such as the differential total electron content.

This project was conducted during the 2019 NASA Frontier Development Lab (FDL) program, a public-private partnership between NASA, the SETI Institute, and commercial partners. We wish to thank, in particular, NASA, Google, Intel, ElementAI and SRI, for supporting this project.

References

- [1] J Aarons, B Lin, M Mendillo, K Liou, and M Codrescu. Global positioning system phase fluctuations and ultraviolet images from the Polar satellite. *Journal of Geophysical Research: Space Physics*, 105(A3):5201–5213, 2000.
- [2] Lasse BN Clausen and Hannes Nickisch. Automatic classification of auroral images from the Oslo Auroral THEMIS (OATH) data set using machine learning. *Journal of Geophysical Research: Space Physics*, 123(7):5640–5647, 2018.
- [3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [4] Benjamin Hou and Athanasios Vlontzos. ResNetAE-<https://github.com/farrell236/resnetae>. 2019.
- [5] Yaqi Jin, Jøran I Moen, and Wojciech J Miloch. GPS scintillation effects associated with polar cap patches and substorm auroral activity: Direct comparison. *Journal of Space Weather and Space Climate*, 4:A23, 2014.
- [6] Yaqi Jin, Jøran I Moen, Wojciech J Miloch, Lasse BN Clausen, and Kjellmar Oksavik. Statistical study of the GNSS phase scintillation associated with two types of auroral blobs. *Journal of Geophysical Research: Space Physics*, 121(5):4679–4697, 2016.
- [7] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.
- [8] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.
- [9] SB Mende, SE Harris, HU Frey, V Angelopoulos, CT Russell, E Donovan, B Jackel, M Greffen, and LM Peticolas. The THEMIS array of ground-based observatories for the study of auroral substorms. In *The THEMIS Mission*, pages 357–387. Springer, 2009.
- [10] Kjellmar Oksavik, Christer van der Meeren, Dag Arne Lorentzen, LJ Baddeley, and J Moen. Scintillation and loss of signal lock from poleward moving auroral forms in the cusp ionosphere. *Journal of Geophysical Research: Space Physics*, 120(10):9161–9175, 2015.
- [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [12] L Spogli, Lu Alfonsi, G De Franceschi, V Romano, MHO Aquino, A Dodson, et al. Climatology of GPS ionospheric scintillations over high and mid-latitude European regions. In *Annales Geophysicae*. EGU, 2009.
- [13] Christer van der Meeren, Kjellmar Oksavik, Dag A Lorentzen, Michael T Rietveld, and Lasse BN Clausen. Severe and localized GNSS scintillation at the poleward edge of the nightside auroral oval during intense substorm aurora. *Journal of Geophysical Research: Space Physics*, 120(12):10–607, 2015.
- [14] Qiuju Yang, Dacheng Tao, Desheng Han, and Jimin Liang. Extracting auroral key local structures from all-sky auroral images by artificial intelligence technique. *Journal of Geophysical Research: Space Physics*, 2019.
- [15] Xi Yang, Nannan Wang, Bin Song, and Xinbo Gao. BoSR: A CNN-based aurora image retrieval method. *Neural Networks*, 116:188–197, 2019.