Identifying Anomaly Aurora with Computer Vision Models

UC Berkeley College of Computing, Data Science, and Society

Alex Toohey, Tommy Duong, Sabrina Nazarzai

University of California, Berkeley, Space Sciences Laboratory

Introduction

Overview:

Scientists now have a strong understanding of aurora produced by charged particle precipitation; however, anomalous aurora like STEVE and picket fence (Figure 1 below) are generated by different forces scientists are still working to understand. We developed a computer vision model to automatically identify these anomaly aurora, so that scientists can further study these anomalies to gain a better understanding of how they are generated.

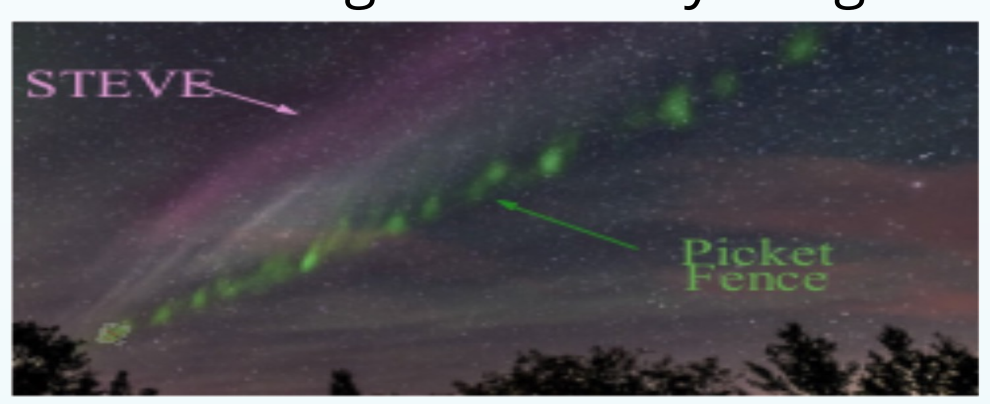

Figure 1: STEVE and the picket fence [1]

Data:

- We utilized images taken every 3 seconds by Transition Region Explorer (TREx) RGB All Sky Imagers (ASI) stationed at 7 locations across Canada
  - This image data was retrieved via the UCalgary AuroraX Data Platform [2]

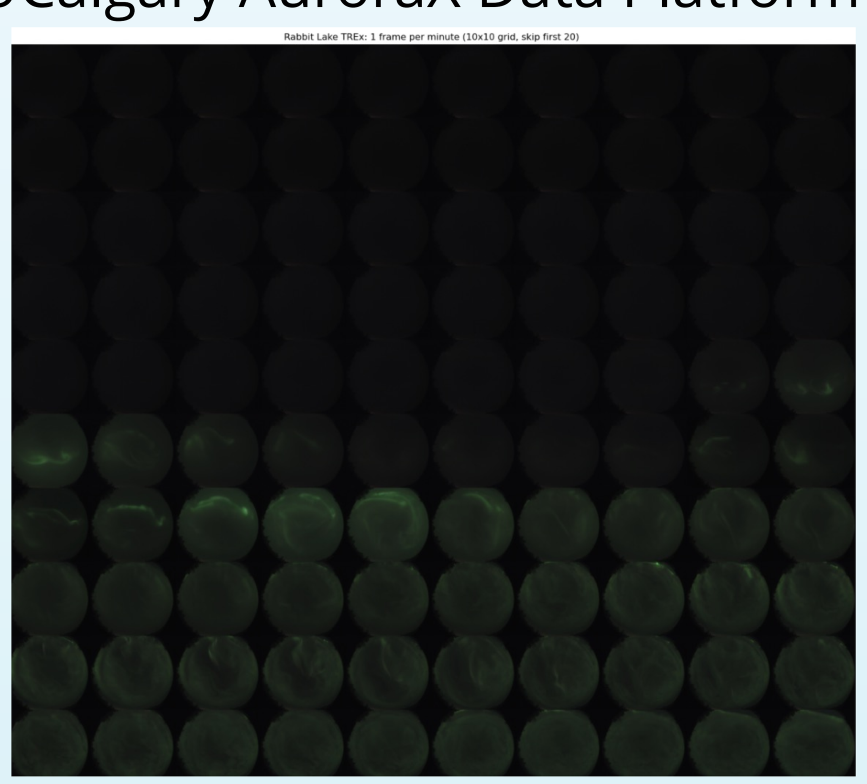

Figure 2: 100 images from our TREx RGB ASI dataset [2]

Methodology

Transfer Learning Approach:

Limited Manually Labelled Dataset: 230 images (only 30 labelled anomalies)

80/20 Train/Test Split

Using foundational models isolates embedding quality as the key differentiator.

Model Comparison:

Baselines: CNN Baseline, CLIP Zero-Shot

Foundational Models: DINOv2 + MLP, DINOv3 + MLP, CLIP + MLP

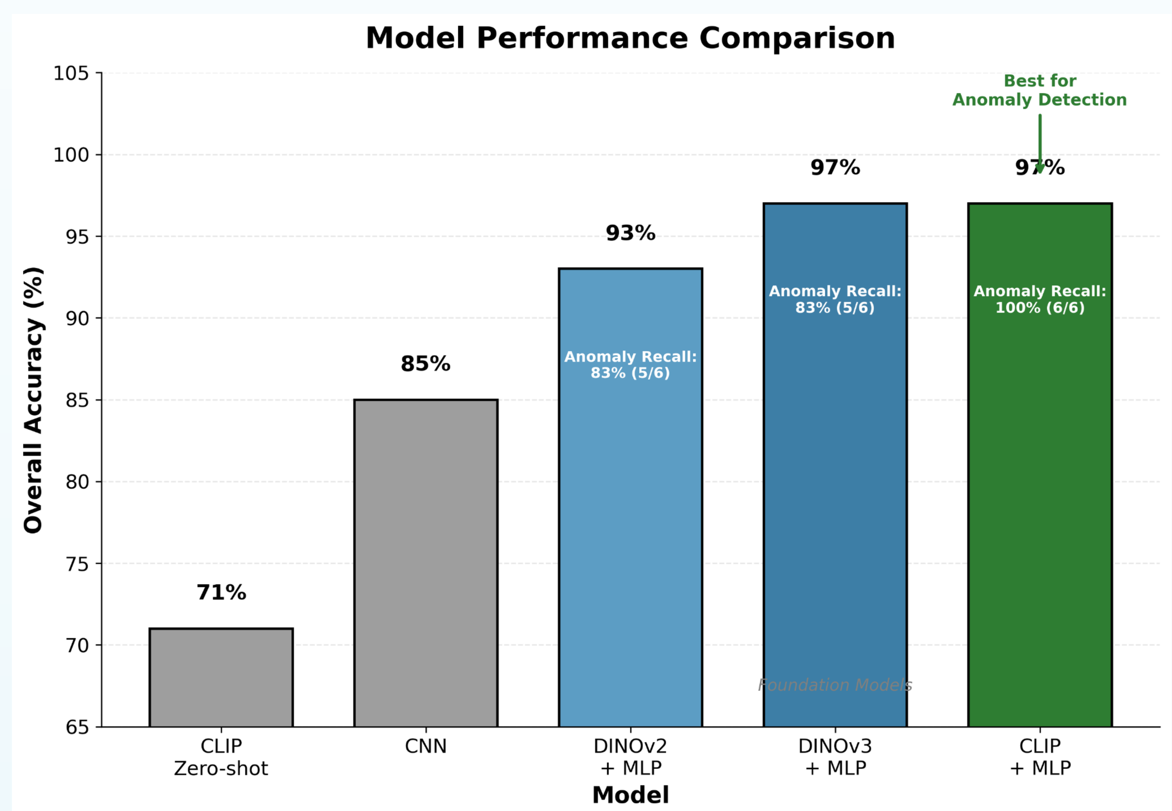

Figure 3. Benchmarking transfer learning approaches for aurora classification. Foundation models with frozen backbones and MLP classifier heads outperform CNN baselines. CLIP's contrastive training objective yields superior anomaly detection (100% test recall) despite matching DINOv3's overall accuracy.

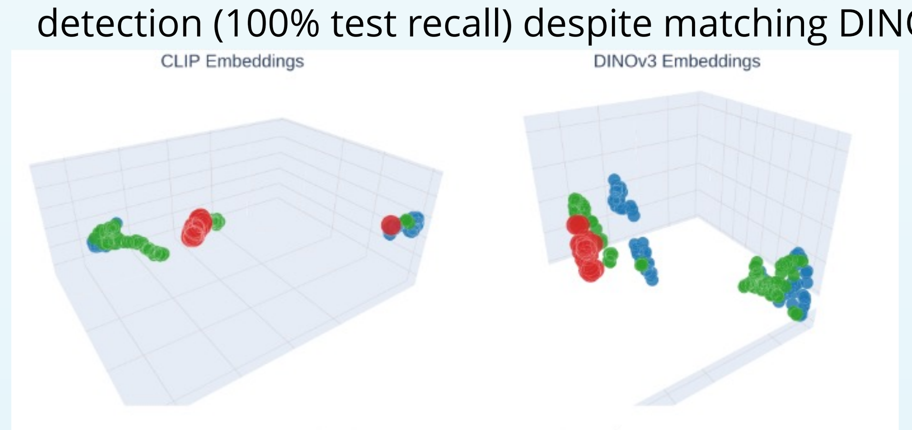

Figure 4. CLIP's contrastive training objective yields superior anomaly detection (100% test recall) despite matching DINOv3's overall accuracy, explained by the tighter clustering.

Model Selection:

- CLIP was chosen as our final embedding model as its contrastive, vision-language embeddings create clear semantic boundaries across classes leading to greater distances between distinct embeddings (Figure 4)
- CLIP was therefore able to detect key image differences necessary to identify anomaly aurora, with the highest anomaly recall out of all models we evaluated (Figure 3)

Results

Model Evaluation:

A stratified evaluation across predicted classes and confidence levels from a diverse set of 150 2023 Rabbit Lake images showed that the model is highly sensitive to auroral activity and especially effective at detecting rare anomaly events. Precision and recall patterns highlight this behavior:

- No Aurora (98 percent precision, 59 percent recall)
- Aurora (41 percent precision, 77 percent recall)
- Anomaly (72 percent precision, 82 percent recall).

The stratified design also revealed that bright sky conditions such as moonlight and clouds common in this diverse test set are the main source of confusion.

Human in the Loop and Batch Insights:

A targeted human in the loop workflow corrected predictions under 55 percent confidence and improved the training dataset by exposing consistent bright-sky errors. Large scale batch classification of more than 120,000 images confirmed the model's high-recall behavior, surfacing many anomaly candidates and ensuring that true anomalies are rarely missed even if some clouds appear as false positives.

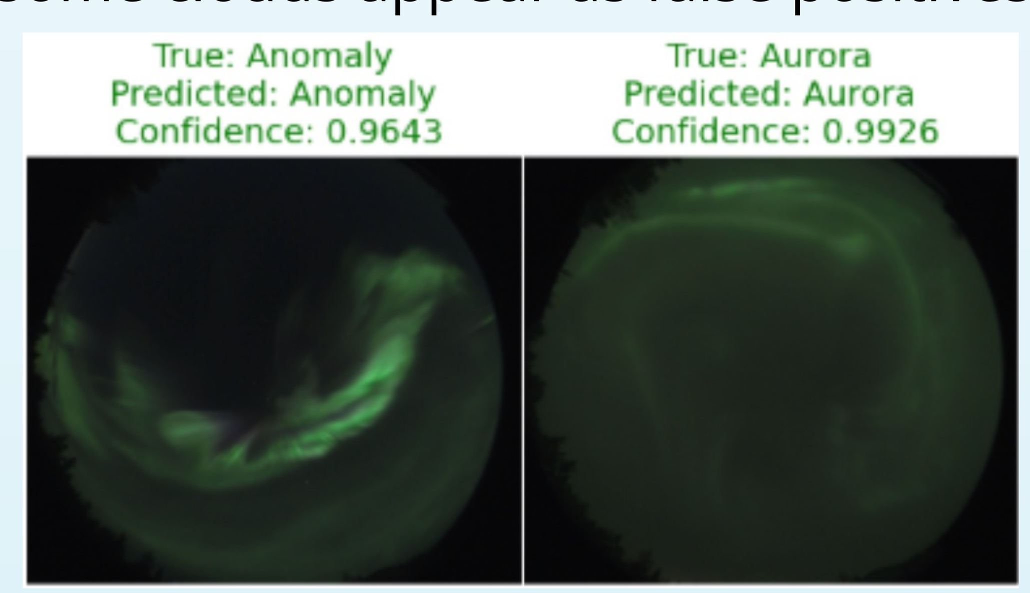

Figure 5. Comparison between images our model classified as anomaly (left) and normal aurora (right)

Conclusion

Advancing Anomaly Aurora Research:

- Our classification pipeline identified hundreds of anomaly aurora instances for scientists to investigate further

Key Insights:

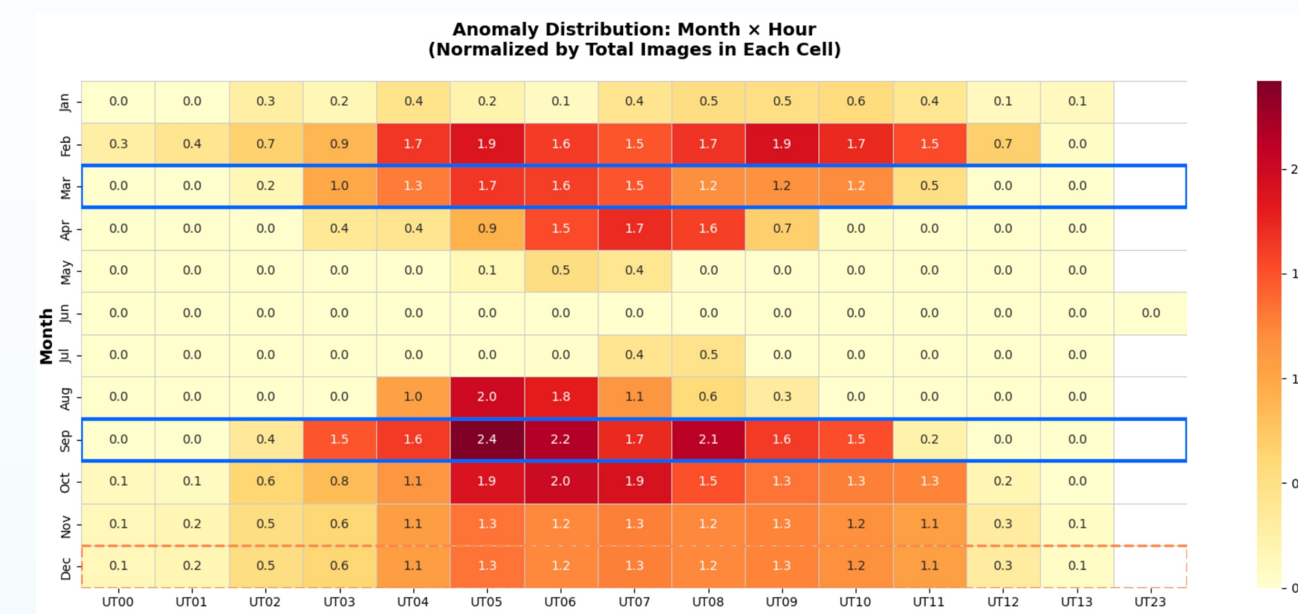

Figure 6. Heatmap of Normalized Anomaly Percentages Across Month & Hour

- Our model detected a much higher percentage of anomalies around the Equinoxes (March & September) and around the middle of the night

Limitations & Future Improvements:

- CLIP-based computer vision model still struggles with classifying images that are highly cloudy or have a bright moon
  - Training data could be made more robust with more cloud-covered images and image augmentations
- More diverse anomaly instances are needed in the training set to detect a higher variety of rare anomalies
- Including temporal component in model could also improve performance

References

[1] Nishimura, Y., Dyer, A., Kangas, L., Donovan, E., & Angelopoulos, V. (2023). Unsolved problems in Strong Thermal Emission Velocity Enhancement (STEVE) and the picket fence. Frontiers in Astronomy and Space Sciences, 10, 1087974.

[2] University of Calgary Auroral Imaging Group. (2025). Aurora Data Portal and API Documentation. Retrieved from https://aurora.phys.ucalgary.ca/

Acknowledgements

We would like to thank Brian Harding and Claire Gasque for all their mentorship and support throughout this project, as well as the CDSS Data Discovery Program for making this project possible.