Deep Learning-Based Classification of GEOs using Unresolved Spectral Data

George V. Landon
Engineering and Computer Science, Cedarville University
David Strong
Strong EO Imaging, Inc.

Timothy Giblin
i2 Strategic Services, LLC

Benjamin Roth
Department of Physics and Meteorology, USAF Academy

Francis Chun
Department of Physics and Meteorology, USAF Academy

ABSTRACT

With the continuously increasing number of objects in orbit, additional methods utilizing fast and broadly deployed sensors to maintain Space Domain Awareness are critical. However, densely deploying sensors that provide spatially resolved imaging capabilities for Geostationary Earth Orbit (GEO) satellites is prohibitively resource-intensive. Therefore, to achieve densely deployed ground-based sensors, a reduction in resolving capabilities is required, which in turn limits the use of classic image-based identification and classification methods. This work explores state-of-the-art deep learning classification architectures that train on non-spatially resolved spectral satellite observations. Leveraging satellite spectral observations in raw spectral flux allows the development of a classification model that can be used for satellite characterization through a simplified preprocessing pipeline. To support methodologies developed in this work, different methods of satellite characterization are performed on five evenings of data, from the United States Air Force Academy Falcon Telescope Network, where non-resolved slitless spectroscopy observations of satellites are automatically classified by satellite, bus, and manufacturer.

While convolutional neural networks (CNNs) have driven significant improvements in classification results over the past several years in many computer vision tasks, the recent development of transformers has put new focus on what is possible in object classification tasks. Both Vision Transformers (ViTs) and Shifted-window (Swin) Transformers have provided exceptional improvements in classification and have led to the more recent improvements of CNNs in the development of ConvNeXt, which is an architecture that seeks to produce results that are similar to, or even outperform ViTs and Swin Transformers while maintaining the simplicity of classic CNNs. This work develops a 1D ConvNeXt architecture and demonstrates its efficiency in satellite classification using slitless spectroscopy observations while developing support for choices of other important training properties such as loss function and optimizer.

Moreover, data preparation remains a critically important step in both architecture design and training. While the observations recorded spectroscopy values from 400-800nm, extracting common bandpass values typical of Kron/Cousins filters provides a method for dimensionality reduction without reducing classification accuracy. To build a robust satellite classification system for Space Domain Awareness, training and validation datasets are carefully separated to mimic real-world scenarios. In this work, training is never performed on "future" or "concurrent" observations of satellites. In an effort to avoid the unfair advantage of learning atmospheric conditions or other conditions that may vary temporally, validation is always performed on observations that occur after observation dates used for training.

The 1D ConvNeXt architecture is deployed on a large dataset of slitless spectroscopy observations of 20 satellites over five evenings. The results explore classification accuracy when using calendar dates later than the training dates. This work seeks to evaluate whether models can accurately classify satellites from future, unseen observations. Results attempt to address which satellite characteristics, such as bus, configuration, or platform are the most robust for classification accuracy of future observations. Bus classification accuracy appears the most consistent in performance, reaching 80% accuracy.

1. INTRODUCTION

The rapid growth in the number of space objects (e.g satellites and space debris) has created increasing challenges in efforts to detect, track, identify, and characterize them. Space situational awareness (SSA) is critical to the United States Space Force and Air Force while space congestion requires new methods for tracking and identifying objects in this new warfighting domain.

Satellites in low-earth orbit (LEO) and geosynchronous-earth orbit (GEO) are of particular interest due both to the importance of these regions and the added density of objects in these areas of spaces. Moreover, optical telescopes remain a standard method of surveillance for objects in these regions and spatially resolved imaging of satellites offer direct means of identification and classification. However, when considering the exponential growth of space objects, the network of spatial resolved imaging telescopes is not adequate in their number or availability. Moreover, even these telescopes are not adequate in resolving all satellites in GEO due to their small size relative to distance from the sensors.

Considering the critical need for SSA, it is necessary to employ telescopes without resolved imaging capabilities. This work proposes to develop new satellite characterization techniques using non- spatially resolved imaging. These telescopes acquire satellite data using photometry, spectroscopy, and polarimetry. Each of the sensor modalities provide information that may individually be used for satellite identification and characterization. This work will explore both uni-modal and multi-modal datasets to develop neural networks that can classify satellites based on these sensing modalities while focusing on spectroscopy observations.

2. RELATED WORK

Rapid development in deep learning architectures has led to numerous opportunities to develop new methods for machine learning. From the first introduction of the transformer and attention[8], classification techniques based on deep learning techniques have rapidly improved. Moving from Natural Language Processing (NLP), two parallel techniques were developed specifically for vision-based classification tasks on images. The Vision Transformer[3] and the Swin Transformer[4]. With the popularity of transformers reaching all areas of machine learning, the classic approaches were still improved and benchmarked against them. Building on transformer successes while exploring classic convolution-based neural networks, the ConvNext architecture[5] was developed and demonstrated to compete with and even outperform transformer-based networks.

These rapid advancements in NLP and Computer Vision tasks have led to exploration and adoption in SDA methods. Initial work [1] comparing various machine learning models has classified satellites by bus-type using tracks, time-series observations as a non-geosynchronous satellite crosses a telescope's FOV. Other work has used slitless spectroscopy [7] of geosynchronous satellites (GEOs) to classify satellites based on spectroscopy measurements of 20 different GEO satellites [11]. Others have recognized the limited nature of many SDA datasets and applied initial training with larger star catalogs and fine-tuning for SDA applications [2].

It is common practice to take an entire dataset and randomly divide between training and validation datasets. Whether it is from the original observations [7] or augmented observations [2], this makes validation and test performance overly optimistic since it is easier to interpolate within the same distribution. In this work, a specific effort is made to separate training data and validation data by time of acquisition: a temporal split. To mimic real-world SDA applications, training is performed on older observations and validation is performed on new observations which require a model to extrapolate to unseen conditions to accurately classify GEOs.

3. MODEL ARCHITECTURE

While the development of the transformer led to a shift in deep learning development, parallel work has demonstrated that convolution neural networks can still outperform transformers in many contexts. ConvNeXt has specifically demonstrated performance outpacing transformers and leads to the architecture presented here. While the original ConvNext architecture was designed for 2D image inputs, for spectral data from unresolved satellites, only 1D inputs are available. Therefore the architecture is modified from 2D to 1D inputs.

3.1 Convolutional block structure

As shown in Fig. 1, the ConvNeXt Block combines ResNet-style convolution while integrating vision transformer (ViT) ideas like spatial attention. First in the block, a Depthwise convolution acts like spatial attention followed by a LayerNorm and two PointWise convolutions to work like the multi-layered perceptrons in ViTs.

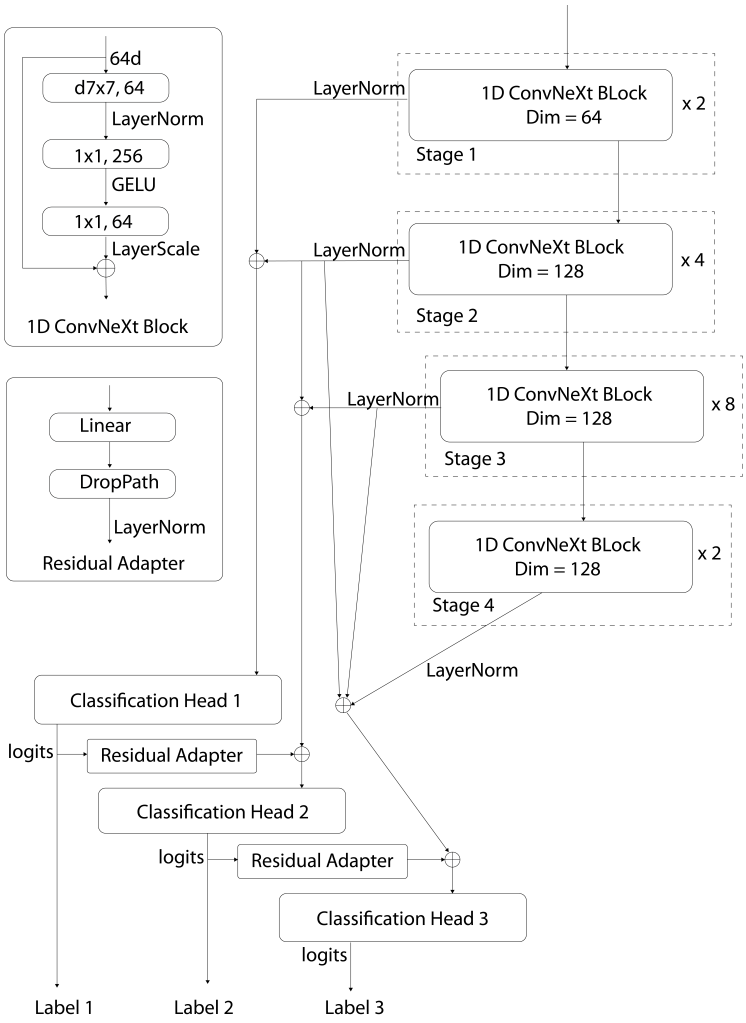


Fig. 1: The proposed 1D ConvNeXt-based architecture for unresolved spectral classification. The backbone consists of stacked ConvNeXt blocks organized into four stages, progressively increasing feature dimensionality. Residual adapters connect the outputs of each classification head back into the shared feature space, enabling hierarchical multi-task learning. The model simultaneously predicts bus family (Label 1), bus variant (Label 2), and platform (Label 3), with earlier, broader predictions informing the more specific classifications.

As these ConvNeXt blocks are stacked to build the network, the model is able to learn low-level flux features up to higher-level features related to bus and platform attributes. The stacking and residual connections are shown in Fig. 1.

3.2 Residual additive conditioning

To handle the multiple classification tasks simultaneously, multiple network heads are chained together to implement residual additive conditioning. Hierarchical Multi-task Classification[9] has been introduced to handle labels that belong to classes that are hierarchically structured. To improve on this, residual connections are added between each head in a chain to allow for backpropagation during training. In this work, three simultaneous classification tasks are attempted: Bus Family (head 1), Bus Variant (head 2), and Platform (head 3).

Each classification head is linked in sequence using residual adapters. After one head produces logits, a small linear

adapter projects those logits back into the feature space of the shared backbone. This projection is added to the features passed to the next head, forming an additive residual connection (shown in the bottom of fig. 1). DropPath randomly drops residual updates to improve robustness. LayerNorm ensures stability after each residual addition.

This design is inspired by residual adapters in multi-task learning [6], but adapted here for hierarchical satellite classification. Each head not only predicts its own class (e.g. bus family) but also provides conditioning information to downstream heads (e.g. bus variant, platform). Consequently, later heads benefit from the hierarchical structure of the problem. Residual adapters act as information bridges between tasks, allowing classification heads to share signal without requiring joint optimization at every step.

3.3 Masked Autoencoder Pretraining

Recently introduced to improve accuracy in an updated version of ConvNext, ConvNext version 2 [10], a self-supervised pre-training step was employed. Using masked autoencoders, the backbone network was pretrained to avoid gradient issues during early supervised training stages. This work employs a 75% mask of flux data as a pre-training step on the backbone only for autoencoding.

3.4 Curriculum Training

During training, an outer loop serves as a global curriculum manager. To improve stability and allow the network to specialize progressively, training was organized into curriculum phases. The curriculum manager cycles through a sequence of training inner loops where different classification heads are emphasized while the backbone is either frozen or unfrozen.

The curriculum begins with a joint warm-up phase in which all three heads (Bus Family, Bus Variant, and Platform) are trained simultaneously. During this phase, the backbone remains unfrozen and the residual adapters are enabled, allowing gradients from all heads to condition shared representations. This encourages the model to learn broad spectral features that support all three tasks.

After the warm-up, the curriculum transitions to specialization phases. Each head is trained independently while the backbone is frozen, ensuring that improvements are isolated to the classifier layers and their adapters. First, the bus family head is trained with a weighted combination of Balanced Softmax and cross-entropy losses. Second, bus variant head is trained with Balanced Softmax and cross-entropy loss with label smoothing. Third, platform head is trained with cross-entropy loss with label smoothing. By freezing the backbone during these specialization phases, the shared feature extractor remains stable while each head adapts to its specific task distribution. This is specified in Table 1.

The global curriculum manager repeats this sequence multiple times (ten in this work), with linearly decaying learning rates between iterations. This staged progression allows the network to first learn generalizable spectral embeddings, then refine each classification task individually.

Head Index	Phase	Backbone Status
$1, 2, 3$	Joint warm-up	Unfrozen
1	Bus Family specialization	Frozen (except Bus head)
2	Bus Variant specialization	Frozen (except Variant head)
3	Platform specialization	Frozen (except Sat head)

Table 1: Curriculum training phases used in the Global Curriculum Manager. Training begins with a joint warm-up phase where all three classification heads are updated together while the backbone remains unfrozen. Subsequent specialization phases focus on individual heads (bus family, bus variant, and platform) while the backbone is frozen, ensuring stable feature extraction and head-specific refinement.

3.5 Loss functions

To handle both the residual aspect of the chained heads as well as the unbalanced labels in multiple heads, multiple loss functions are used to benefit training. Head 1 (bus families) has the fewest classes and cross entropy loss with label smoothing and class weights is used. For heads 2 and 3, they are both dependent on previous classifications within the additive conditioning.

First, Balanced Softmax Loss is a variant of cross-entropy that corrects for label imbalance by adding the log of the per-class sample counts to the logits. This loss function helps prevent bias toward majority classes when training with

highly imbalanced data and is applied to the bus family and variant heads. Standard Cross Entropy Loss with label smoothing is applied on the platform head to prevent the model from becoming too confident. An additional loss function is calculated on each head based on F1 score to account for both precision and recall. Specifically, a Soft F1 Loss is used as a differentiable approximation of F1 score.

For the warm-up phase, all three heads are trained together and the total loss is a weighted sum, where bus family, bus variant, and platform sat are with similar weights.

For single-head training, the respective head is strongly weighted (2x), while the others are downweighted (0.05×). This allows the model to specialize per head but still propagate gradients through the chain.

Head	**Primary Loss**
Bus Family (Head 1)	Balanced Softmax Loss + Soft F1 Loss
Bus Variant (Head 2)	Balanced Softmax Loss + Soft F1 Loss
Platform / Satellite (Head 3)	Cross Entropy Loss + Soft F1 Loss
Joint (all heads)	Weighted Combination of above losses

Table 2: Primary loss functions applied to each classification head. Bus family and bus variant are trained with Balanced Softmax Loss and Soft F1 Loss to address class imbalance and optimize both precision and recall. Platform-level classification uses Cross Entropy with Soft F1 Loss to prevent overconfidence. During joint training, a weighted combination of all three losses ensures stable multi-task learning across heads.

4. DATASET

Classifier			Observation Date				
Platform	Bus Family	Bus Variant	2022-08-11	2022-08-13	2022-08-24	2022-08-25	2022-08-28
AMC15	A2100AX	A2100AXS	5	5	5	10	5
ANIKF1R	EUROSTAR-3000	EUROSTAR-3000S	5	5	5	10	5
ANIKF2	BSS-702HP	BSS-702HP	4	5	5	10	5
ANIKG1	SSL-1300	SSL-1300	5	5	5	10	5
DTV10	BSS-702HP	BSS-702HP	5	5	5	10	5
DTV12	BSS-702HP	BSS-702HP	5	5	5	10	5
DTV14	SSL-1300	SSL-1300	5	5	5	10	5
DTV15	EUROSTAR-3000	EUROSTAR-3000	5	5	5	10	5
ECHOSTAR10	A2100AX	A2100AXS	5	5	5	10	5
ECHOSTAR11	SSL-1300	SSL-1300	5	5	5	10	5
ECHOSTAR17	SSL-1300	SSL-1300	5	5	5	10	5
GALAXY16	SSL-1300	SSL-1300	5	5	5	10	5
INMARSAT4F3	EUROSTAR-3000	EUROSTAR-3000GM	5	5	5	10	5
MEXSAT3	STAR-2.4BUS	STAR-2.4BUS	0	5	5	10	5
NIMIQ2	A2100AX	A2100AX	5	5	5	10	5
SES11	EUROSTAR-3000	EUROSTAR-3000	5	5	5	10	5
SES3	STAR-2.4BUS	STAR-2.4BUS	5	5	5	10	5
SKYTERRA1	BSS-702HP	BSS-702HP-GEM	5	5	5	10	5
SPACEWAY3	BSS-702HP	BSS-702HP	5	5	5	10	5
WILDBLUE1	SSL-1300	SSL-1300	5	5	5	10	5
Dataset Split			Training				Validation

Table 3: Observation schedule for the 20 satellites in the SURi 2 dataset, showing bus family, bus variant, and platform labels. Each satellite was observed multiple times over five nights in August 2022. Data from August 11–25 was used for training, while August 28 was reserved for validation, creating a temporal split to test generalization to future, unseen observations.

To test the 1D ConvNeXt architecture for classifying satellites by bus family, variant and platform on unseen observation dates, this work uses the SURi 2 dataset. This dataset contains 20 GEO satellites where spectral measurements

from 380-880nm were recorded for each observation using slitless spectroscopy. Additional information regarding optics, positioning, and additional recording details are provided by Yee et al. [11].

Table 3 shows the dataset that was used in this work from the SURI 2 dataset. In this table, it can be seen that 20 different satellites were typically observed at least 5 times each on 5 different evenings of August 2022.

In the bottom row of Table 3, it is shown how datasets are split into Training and Validation. This provides a clear delimiter between historical (2022-08-11 - 2022-08-25) and future (2022-08-28) observations where training will not be performed on any observations that occur during or after the validation observations.

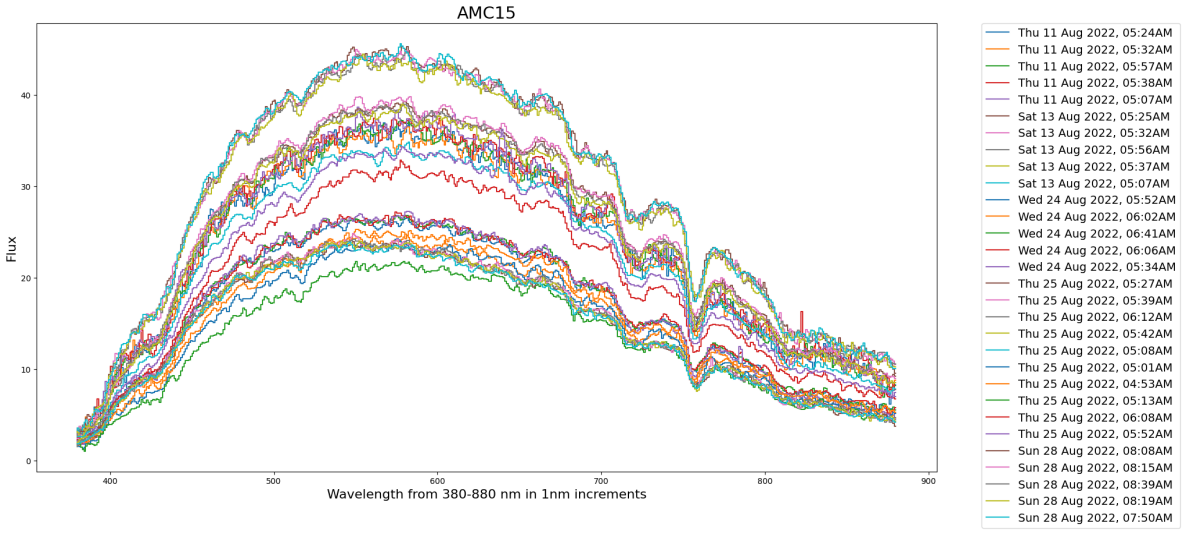


Fig. 2: Raw slitless spectroscopy flux measurements of the AMC15 satellite collected across multiple observation dates between August 11 and August 28, 2022. Each curve represents a single observation of spectral flux across the 380–880 nm wavelength range in 1 nm increments. AMC15 belongs to the A2100AX bus family and specifically to the A2100AXS bus variant. The repeated measurements illustrate temporal variability due to atmospheric and observational conditions while maintaining the overall spectral shape characteristic of this bus design. These raw flux curves form the input data used in preprocessing and subsequent classification.

Fig. 2 and Fig. 4 show all spectral recordings for the AMC15 satellite and the Echostar 11 satellite, respectively. While this dataset contains additional processed versions of each observations (e.g. solar analog scaled, etc...), to maintain the simplest pipeline for SDA applications, only the preprocessed flux is used.

To process the data for input into the classification network, only a subset of spectral values are used. Following the Visual filter with some lower coverage of the Red filter of the Kron/Cousins specification, a 200nm band from 500nm-700nm is extracted from the complete 380nm-880nm range. This flux channel is then scaled using 1% - 99% quantiles of the 200nm band and set as the first input channel. The second input channel contains positional encoding. These wavelength values, “where in the spectrum”, are normalized 0..1. Therefore, the dataset is arranged such that the input is separated into 2 channels as seen in Fig. 3 and Fig. 5 where both the x and y values are stored as the input dimensions.

Moreover, each of these processed spectral observations is labeled with three labels: Bus Family Classification, Bus Variant Classification, and Platform Classification.

5. RESULTS

Data are temporally split into training observations (2022-08-11 - 2022-08-25) and validation observations (2022-08-28). Training is performed first as Mask Autoencoder (MAE) pretraining step on the flux channel of the network. Then training is performed on each head simultaneously with an equally weighted loss for each classification head.

The classification accuracy for the validation dataset shows a strong performance for the Satellite Bus Family of 80%.

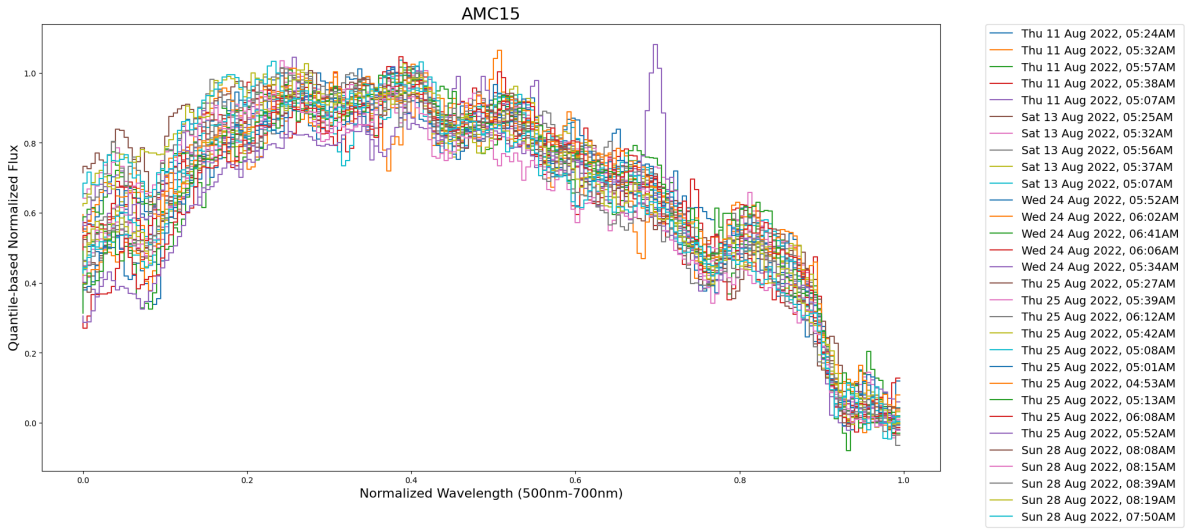


Fig. 3: Quantile-based normalized flux for AMC15 (shown in Fig. 2). The 500–700 nm band is extracted from the raw spectra and scaled using the 1%–99% quantiles, with wavelength positions normalized from 0 to 1. This preprocessing reduces variability across observations while preserving spectral shape for input into the 1D ConvNeXt model.

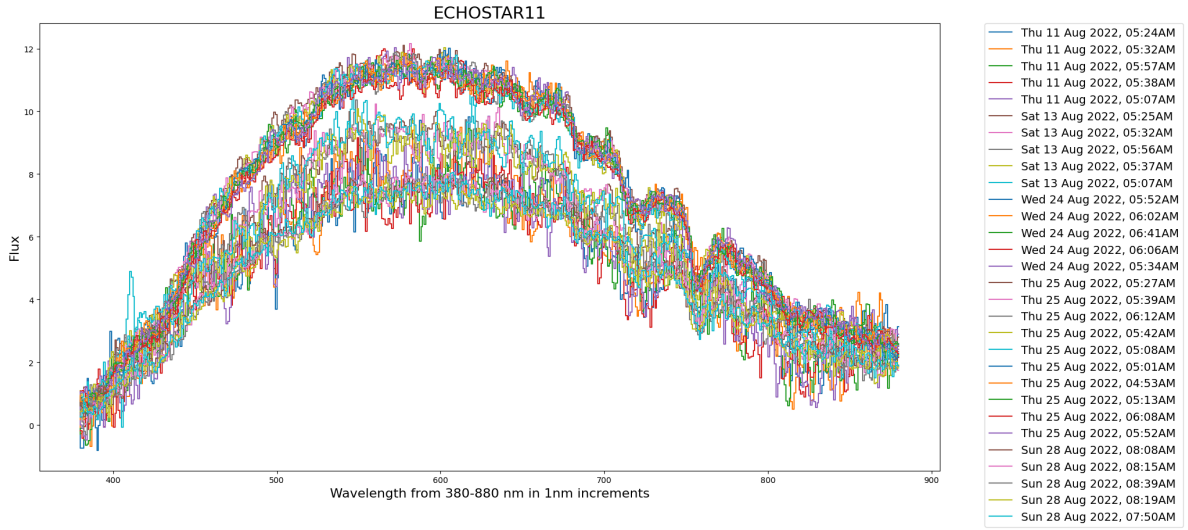


Fig. 4: Raw slitless spectroscopy flux measurements of the EchoStar11 satellite collected across multiple observation dates between August 11 and August 28, 2022. Each curve represents a single observation of spectral flux across the 380–880 nm wavelength range in 1 nm increments. EchoStar11 belongs to the SSL-1300 bus family and is classified within the SSL-1300 bus variant due to limited knowledge on the SSL-1300 variants. The repeated measurements illustrate temporal variability due to atmospheric and observational conditions while maintaining the overall spectral shape characteristic of this bus design. These raw flux curves form the input data used in preprocessing and subsequent classification.

Each class has majority classification correct while there is some cross labeling between BSS-702HP and SSL-1300 which may be due to the amount of satellites in these families or even the number of variants in each family.

As the classification task becomes more specific, the model's performance shows a noticeable decline in accuracy. While bus family classification achieves robust results, neither the bus variant classifier (Head 2) nor the platform classifier (Head 3) is able to exceed 50% accuracy on the validation dataset. This reduction highlights a key limitation of the temporal split strategy: unlike random splits, which allow models to interpolate within familiar distributions, a

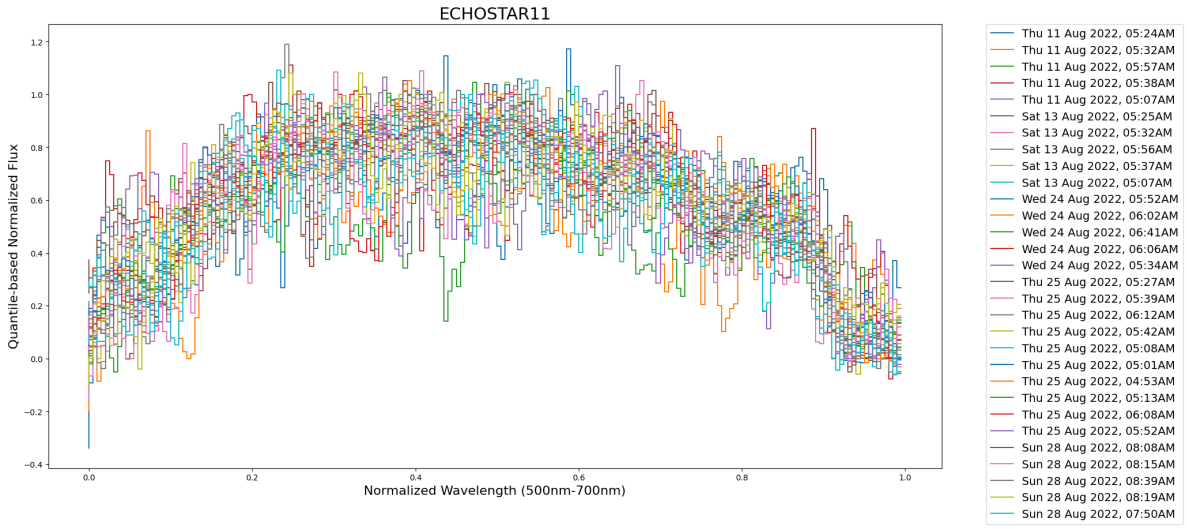


Fig. 5: Quantile-based normalized flux for EchoStar11 (shown in Fig. 4). The 500–700 nm band is extracted from the raw spectra and scaled using the 1%–99% quantiles, with wavelength positions normalized from 0 to 1. This preprocessing reduces observational variability while preserving the spectral shape used as input to the 1D ConvNeXt model.

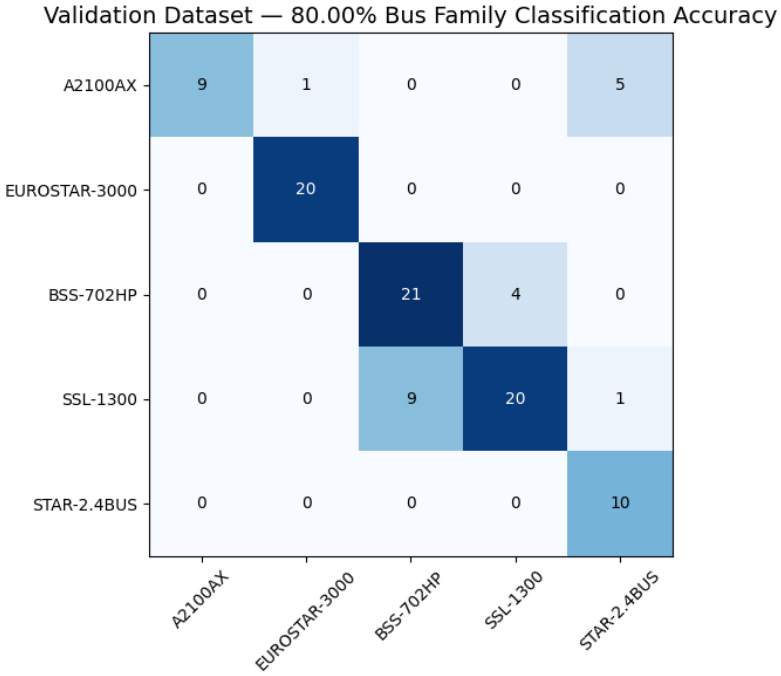


Fig. 6: Confusion matrix for bus family classification on the validation dataset from 2022-08-28 observation, showing an overall accuracy of 80%. Each row represents the true bus family, and each column represents the predicted class. The model correctly classifies the majority of observations for EUROSTAR-3000, BSS-702HP, and STAR-2.4BUS, which appear as strong diagonal entries. Some misclassifications occur between SSL-1300 and BSS-702HP, as well as between A2100AX and other families, reflecting similarities in their spectral features. These results demonstrate that bus family characteristics are robust enough to be consistently identified across unseen observation dates

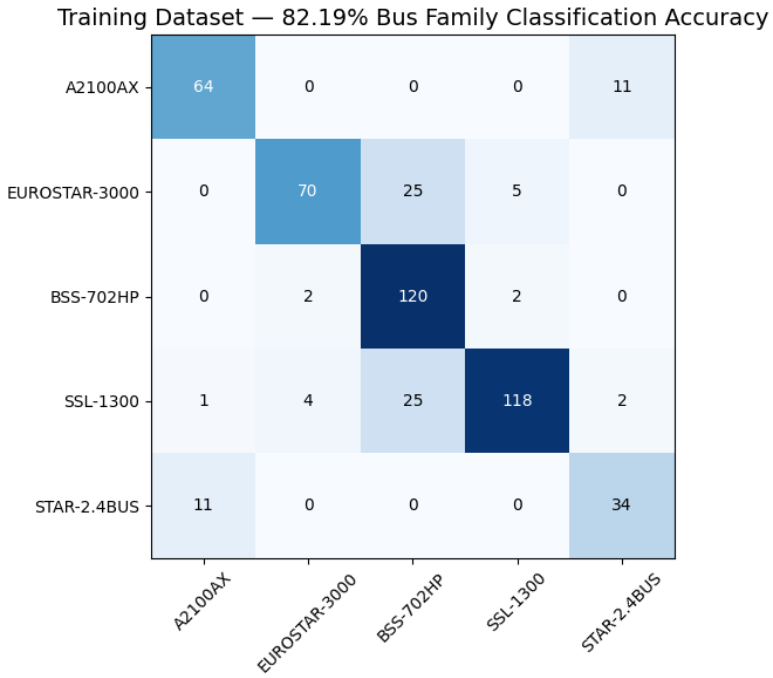


Fig. 7: Confusion matrix for bus family classification on the training dataset from 2022-08-11-2022-08-25 observations, showing an overall accuracy of 82.19%. Correct classifications dominate the diagonal entries, especially for BSS-702HP and SSL-1300, which have the highest number of correctly identified samples. Some misclassifications are observed between EUROSTAR-3000 and both BSS-702HP and SSL-1300, reflecting overlap in their spectral signatures. A smaller degree of confusion also occurs between A2100AX and STAR-2.4BUS. These results confirm that the model performs strongly on training data, but comparison with the validation results (Fig. 6) highlights the importance of testing under temporal splits to ensure generalization beyond known observations.

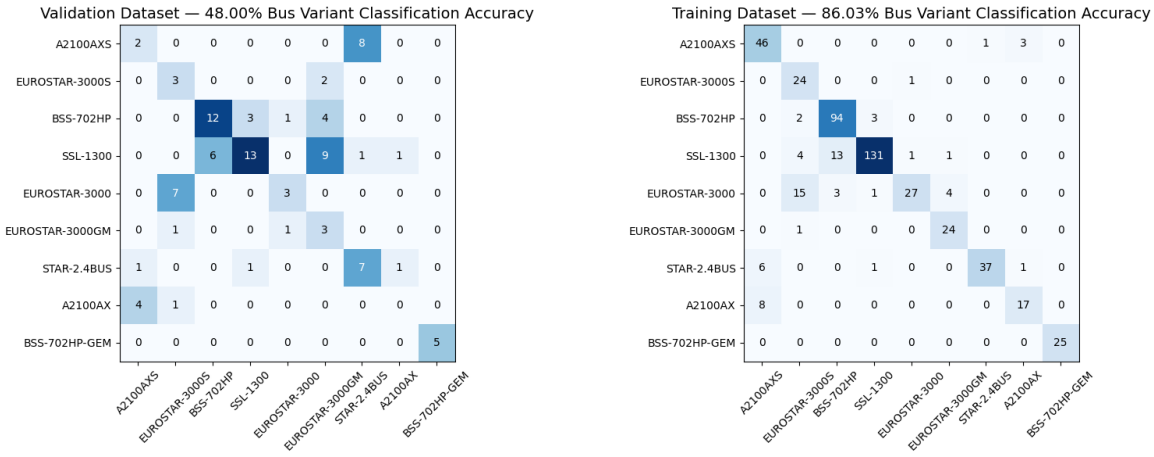
temporal split requires the model to extrapolate to entirely new observational conditions. Consequently, the classifier must contend not only with spectral variability introduced by atmospheric or instrumental factors, but also with the inherent challenge of identifying finer distinctions between similar classes in data it has never encountered before.

For the bus variant task, the model achieves an accuracy of 48%, which reflects the difficulty in distinguishing between subclasses within the same bus family. While some of this limitation can be attributed to class imbalance, with certain bus variants heavily represented while others occur less frequently, the broader issue is that many variants exhibit highly similar spectral signatures. This makes it difficult for the network to identify unique features that reliably separate one variant from another. In practice, this suggests that unresolved spectral data provides a stronger signal for broad categorical distinctions, such as bus family, than for more detailed subclassifications.

At the platform level, classification accuracy drops significantly, with the model achieving only 33% on the validation dataset. The confusion matrix in Fig. 9a illustrates the nature of these errors. While certain satellites, such as ANIKG1, SKYTERRA1, and SES11, are classified with high reliability (with multiple correct identifications concentrated along the diagonal), many others exhibit frequent misclassifications.

For example, AMC15 is consistently misidentified as INMARSAT4F3, confirming the earlier observation that satellites with overlapping spectral profiles are difficult to distinguish. Similarly, several DirectTV platforms (DTV10, DTV12, DTV15) are often confused with each other, reflecting the spectral similarity of satellites built within the same bus family. These patterns suggest that the classifier is able to capture broad differences between bus families but struggles to resolve fine intra-family distinctions at the platform level.

Another feature highlighted by the confusion matrix is the degree of cross-platform mixing. For instance, WILDBLUE1 and ECHOSTAR satellites show scattered misclassifications across unrelated platforms, further reducing overall accuracy. Such errors indicate that when the model cannot identify strong discriminative features, it tends



(a) Validation Dataset from 2022-08-28 observation.

(b) Training Dataset from 2022-08-11-2022-08-25 observations.

Fig. 8: Confusion matrices for bus variant classification on the 8a validation dataset and 8b training dataset. On the training data, the model achieves 86.03% accuracy, with strong diagonal dominance across most bus variants, particularly SSL-1300 and BSS-702HP. However, performance drops significantly on the validation set, with overall accuracy reduced to 48.00%. Misclassifications occur frequently between variants of the same bus family, such as EUROSTAR-3000 vs. EUROSTAR-3000S/GM, reflecting the high spectral similarity of these subclasses. This contrast highlights the difficulty of distinguishing between closely related bus variants from unresolved spectra, even when higher-level family classification remains robust.

to fall back on incorrect predictions spread across multiple candidate classes.

Taken together, these results reinforce the conclusion that unresolved spectroscopy is more effective for high-level classification tasks (such as bus family) than for fine-grained platform identification. The confusion matrix makes clear that while the model can recognize distinctive cases, the lack of unique spectral signatures among many satellites prevents reliable separation at the platform level. Addressing this challenge may require augmenting spectroscopy with additional sensing modalities (e.g., photometry or polarimetry) or incorporating temporal and contextual features to provide further discriminatory power. This work shows strong promise in classifying GEO from past observations and specifically classifying them by their bus family.

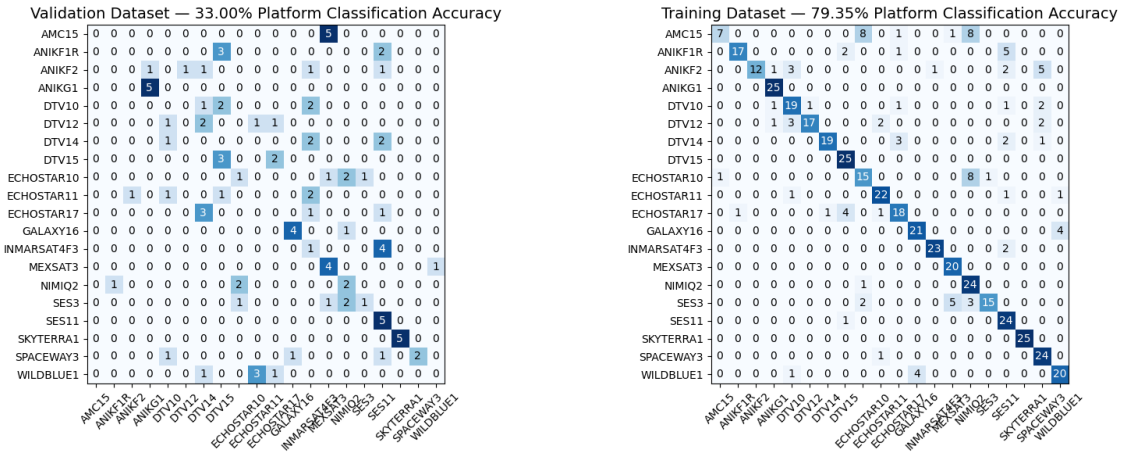
6. CONCLUSION

This work presents a pipeline for GEO classification utilizing recent developments in deep learning architectures for improved Space Domain Awareness. A custom 1D ConvNeXt multi-head classification network was developed and presented taking advantage of progress in transformers and residual CNNs. To replicate real-world SDA applications, a model is trained on observations from the older dates of a dataset and validation is performed on future observations. Results show very strong accuracy in classifying bus-families using only four observation dates and unresolved slitless spectroscopy.

7. REFERENCES

[1] Joshua Collyer Joshua Davis Nicholas Pallecaros Alexander Agathanggelou, Ryan Houghton. Machine learning for satellite characterisation. In *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, 2022.

[2] Leonardo Camacho Damien Eberhardt Chad A. Mello, Matthew Mendoza. Advancing geosynchronous satellite classification utilizing spectral data via fine-tuned pretrained deep learning models. In *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, 2024.



(a) Validation Dataset from 2022-08-28 observation.

(b) Training Dataset from 2022-08-11-2022-08-25 observations.

Fig. 9: Confusion matrices for platform-level classification on the 9a validation dataset and 9b training dataset. On the training data, the model achieves 79.35% accuracy, with most individual satellites correctly identified along the diagonal. However, performance declines sharply on the validation set, with accuracy reduced to 33.00%. Some satellites, such as ANIKG1 and SKYTERRA1, are consistently classified correctly, while others, including AMC15, are frequently misidentified (often as INMARSAT4F3). The contrast between training and validation performance reflects the temporal split strategy, where validation samples are drawn exclusively from later observation dates. This design forces the model to generalize to unseen atmospheric and observational conditions, making fine-grained platform classification especially challenging when spectral differences between satellites are subtle.

[3] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[4] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.

[5] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[6] Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. Learning multiple visual domains with residual adapters. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[7] Roger D. Tippets, Stephen Wakefield, Shannon Young, Ian Ferguson, Christopher Earp-Pitkins, and Francis K. Chun. Slitless spectroscopy of geosynchronous satellites. *Optical Engineering*, 54(10):104103, 2015.

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

[9] Jonatas Wehrmann, Ricardo Cerri, and Rodrigo Barros. Hierarchical multi-label classification networks. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5075–5084. PMLR, 10–15 Jul 2018.

[10] Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[11] Xin Yee, Phan Dao, David Strong, Charles Wetterer, Benjamin Roth, and Francis Chun. Machine learning classification of GEOs using spectral data. In *Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS)*, 2023.

8. DISTRIBUTION STATEMENT

9. DISCLAIMER