Transformers for Maneuver Detection from Navigation Data

Matthew D. Popplewell
matthew.popplewell@advancedspace.com
Advanced Space, LLC
Westminster, Colorado, USA

Nathan Parrish Ré
nathan.re@advancedspace.com
Advanced Space, LLC
Westminster, Colorado, USA

Abstract

Transformers, a type of neural network (NN), have revolutionized the field of natural language processing (NLP), serving as the backbone for nearly every prevalent NLP model (ChatGPT, DALL-E, etc.). Recent studies have shown Transformer models also achieve state-of-the-art performance for tasks involving complex sequences of data. This characteristic makes Transformers a natural choice for handling orbit determination data, which is typically a sparse, multivariate sequence involving multiple phenomenologies with uneven time steps. This study explores the use of Transformer models for spacecraft maneuver detection and characterization. Given a time series of navigation data in which a maneuver is present, the model seeks to predict when the maneuver occurred and the maneuver's magnitude. Results show the NN can predict the beginning of finite maneuvers to within tens of minutes. Additionally, the spacecraft need not be cooperative as the model achieves similar performance for maneuvers that occur during ground station coverage gaps. In conjunction with anomaly classification models, this study introduces a crucial component for autonomous pattern-of-life analysis and spacecraft tracking.

CCS Concepts

• **Applied computing** → *Aerospace*.

Keywords

Transformers, Maneuver Detection, Machine Learning, Space Situational Awareness, Orbit Determination

ACM Reference Format:
Matthew D. Popplewell and Nathan Parrish Ré. 2024. Transformers for Maneuver Detection from Navigation Data. In *Proceedings of 4th International Conference on AI-ML Systems, AI in Space Workshop.* ACM, New York, NY, USA, 6 pages. https://doi.org/3703412.3704410

1 Introduction

The increasing complexity and density of space operations necessitates the automation of spacecraft tracking, anomaly characterization, and maneuver identification in the face of limited ground-based resources. Traditionally, spacecraft navigation relies heavily on experienced human navigators to guide classical algorithms towards reasonable solutions. Statistical models and hard-coded logic have attempted to enable autonomous space situational awareness (SSA) operations but these methods often fall short when it comes to the characterization of subtle, complex maneuvers. As space missions continue to grow more complex, operate in new regimes, and rely on autonomous algorithms, there is a critical need for tools that can not only identify space events but also provide operators with insight into spacecraft behavior.

This paper explores the use of state-of-the-art deep learning techniques for autonomous maneuver characterization. Transformer models are trained on tens of thousands of high-fidelity simulated spacecraft trajectories, radar and optical observations, and navigation filter products in a supervised learning fashion. This approach allows the models to fundamentally learn the data that spacecraft operators work with every day. Furthermore, the models learn to be robust to a wide range of orbital regimes, spacecraft properties, and operational scenarios.

The models presented within are tasked with identifying the start time of an unidentified finite maneuver within a large time window of navigation filter products. This task is typically labor intensive, requiring human operators to sift through dense orbit determination data to identify if/when a signal-of-interest occurs. Maneuvers that occur outside of tracking arcs exacerbate the issue, forcing navigators to use their intuition, experience, and creativity to fine tune traditional models until a converged solution is found. This technology aims to enable the automation of spacecraft operations, freeing navigators to focus on more critical, challenging tasks. The algorithms are designed to be readily integrated in an autonomous SSA suite and ingest incoming streams of spacecraft navigation data.

Autonomous operations are even more challenging for onboard applications. Traditional processes and algorithms often require a human to guide them and, as such, are infeasible for onboard deployment. The limited computational capabilities of flight hardware further limit the applicability of traditional tools. Machine learning (ML) offers a promising alternative to enable onboard, autonomous spaceflight. While ML models require extensive computational resources to train, the evaluation of the models is straightforward and requires minimal memory and compute capabilities. Even current spacecraft hardware has the capability to evaluate complex Transformer neural networks onboard [5].

This paper is organized as follows. Section 2 provides an overview of the necessary background on spacecraft navigation and Transformer models. Section 3 outlines the model architecture, data preprocessing, and information flow through the model. Section 4 describes the dataset used for the study, including the details of the simulated trajectories and measurements. Finally, Section 5 presents the results of the study for three unique use cases.

4th International Conference on AI-ML Systems, AI in Space Workshop, October 8, 2024, Baton Rouge, LA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/3703412.3704410

2 Background

2.1 Spacecraft Navigation

Spacecraft navigation, or orbit determination (OD), is the process of estimating a spacecraft's trajectory, commonly represented as position and velocity relative to a celestial body, and other parameters of interest via the collection of astrometric measurements. Commonly, the best estimate of the object's motion is updated through statistical methods such as Kalman filtering. These processes linearize the highly nonlinear OD problem and seek to iteratively refine the filter's estimate of the spacecraft's state by minimizing the differences between the observed and predicted measurements. A Kalman filter has two primary steps:

(1) A prediction step in which the filter's estimate (t_0) of the parameters of interest and their uncertainties are propagated to the current time step (t_1).
(2) An update step in which the filter's estimate is refined using the difference of the prediction step's results and the observed measurement.

Throughout this paper, "prefit residuals" will refer to the difference between the expected and observed measurements prior to the update step. "Postfit residuals" will refer to this difference after the update step. These residuals are commonly used by human operators to evaluate the performance of the navigation filter. Residuals that are normally distributed about zero mean the filter's estimate and dynamical model accurately represent the actual motion of the spacecraft. Residuals with some trend or signal reveal that some error is present.

When the observed motion is reasonably explained by linear estimates, Kalman filters work well. However, orbit determination is a fickle process that relies heavily on humans-in-the-loop. Kalman filters frequently diverge and require the manual tuning of a large number of parameters. Successful spacecraft navigation requires the experienced navigator's intuition to tune filters and, when anomalies occur, creativity to identify and correct the errors. A previous paper by the same authors introduced ML models that identify and classify systematic anomalies in spacecraft OD by inspecting the navigation filter residuals [5]. The models ingest the filter residuals and context about the observers to infer which small forces, such as atmospheric drag, are mismodeled in the filter dynamics.

A crucial process in spacecraft tracking is the identification and estimation of maneuvers. Large impulsive maneuvers often cause the navigation filter to diverge if not correctly addressed by human operators. Finite maneuvers, due to their extended execution time, may appear similar to other small perturbing forces, such as outgassing. In either case, characterization of the executed maneuver is crucial to maintain a continuous trajectory and tracking. For cooperative spacecraft, the spacecraft operators have a reference design to use as a baseline. Knowledge of the maneuver as-executed is still valuable, however, as comparing the reference and as-flown maneuvers may reveal anomalies in the spacecraft hardware and inform future mission decisions. For non-cooperative spacecraft, information on maneuver design and scheduling is often unavailable to a human operator. Therefore, it is critical to be able to rapidly identify and characterize spacecraft maneuvers for maintaining tracking custody, pattern-of-life analysis, and space situational awareness.

2.2 Transformer models

Transformer models are currently the state-of-the-art deep learning architecture for modeling sequential data. Transformers were introduced for natural language processing (NLP) problems [6] and, after revolutionizing the field, were quickly applied to other sequence problems involving image [2] and time series [7] data. Transformer models are sequence-to-sequence models that rely entirely on a self-attention mechanism without any recurrent layers. Inspired by human cognitive attention, the attention mechanism allows models to capture long-range relationships between tokens in the input sequence more efficiently and effectively than previous architectures. In other words, the model learns to "pay attention" to the most important parts of the data. The model achieves this by comparing all elements in a sequence with all other elements, regardless of their position.

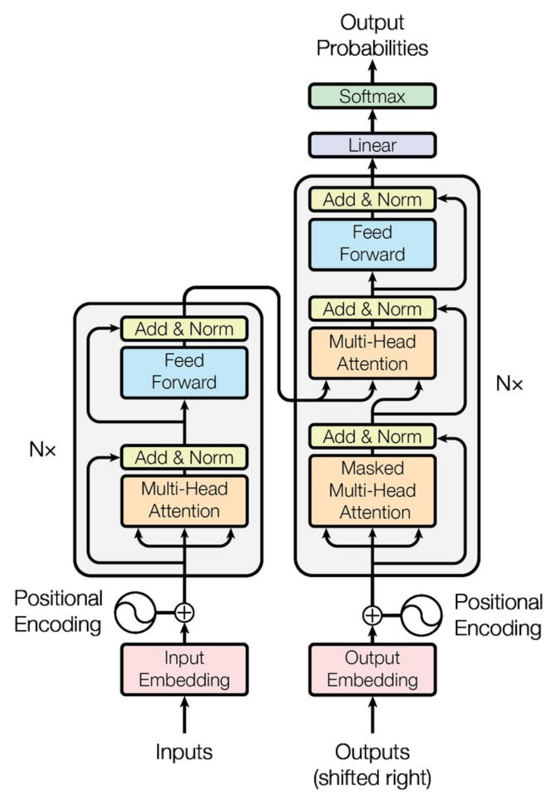

Figure 1: Original Transformer architecture as introduced by Vaswani, et al. [6]

The classic Transformer architecture is shown in Figure 1 and consists of two parts: the encoder, represented by the left side of the image, and the decoder, on the right side of the image. The encoder transforms the input data into an abstract representation that best represents the key features of the sequence. The decoder processes the output of the encoder, and a "target" sequence, to generate an output sequence such as translated text. Both the encoder and decoder consist of multiple attention heads that perform the attention mechanism several times simultaneously, allowing the model to attend to different parts of the sequence. The ability of Transformers to capture multiple relationships in the data makes them well-suited for tasks involving data with complex patterns and signals, such as spacecraft navigation data which is typically a sparse, multivariate time series.

Furthermore, unlike Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) models that are limited to processing sequences one step at a time, Transformer models can process entire sequences simultaneously. This capability reduces training time and leads to the use of larger models. The nature of the attention mechanism also allows transformers to work with long sequence lengths without the vanishing/exploding gradient problems found in RNNs. Whereas RNNs struggle to capture long-term dependencies and context due to vanishing gradients during the backpropagation phase, the self-attention mechanism of Transformers allows them to attend to any part of the input sequence to any other part, regardless of the distance from each other. Consequently, long sets of navigation data can be used effectively, without the computation and model limitations that have limited applications in the past.

3 Model Architecture

The original study of using Transformers for identifying anomalies in OD data used three unique model architectures [5]. This study will present the results of applying the Measurement Transformer (MT) to the characterization of maneuvers. The MT is similar to the BERT architecture [1]. The model contains a Transformer encoder block only and uses a time encoding rather than a positional encoding. Unlike NLP, in which each input token (character or word) can be considered uniformly spaced, spacecraft OD data is riddled with uneven time steps. Commonly, the incoming navigation data will feature dense "tracking passes" while the spacecraft is in-view of a ground station followed by long periods without data while the spacecraft is occulted by a celestial body. A time encoding layer preserves this critical context. The Time2Vec algorithm [4] is used in this study, which provides the model with a vector representation of time.

Additionally, the beginning of each input sequence is prepended with a "class token", first introduced by the Vision Transformer [2]. The class token is a vector of zeros that contains no information about the input sequence, but is still transformed by the NN model. The token is extracted after the final layer of the forward pass and serves as the predicted output of the model. The class token has been shown to be an effective method at reducing a two-dimensional output sequence to a one-dimensional vector prediction.

Data preprocessing follows classical machine learning practices. Most input features are scaled by their corresponding mean and standard deviation to be of approximately magnitude one. The exceptions are ground station ID numbers and time, which are scaled globally such that the same ID number always refers to the same ground station and the time scale is consistent across all samples. Input sets are then padded or trimmed such that all input sequences are of the same length. In the case of padding, vectors of zeroes are appended to the end of the sequence. For sequences that need to be trimmed, indices of the original sequence are randomly sampled to produce the training sample, such that the whole 24 hours of data is still represented.

The data flow of the MT is depicted in Figure 2 and described below:

(1) A Time2Vec layer expands the time element of the feature vector.
(2) A linear layer expands the dimension of each feature vector to the Transformer model's hidden dimension size.
(3) A class token of zeros is appended to the 0th entry of each sequence.
(4) The Transformer encoder layer uses multi-head self attention to draw relationships between different elements of the sequence.
(5) A linear layer reduces the Transformer hidden dimension size to the desired output dimension size, three in the case of predicting the maneuver's start, stop, and magnitude.
(6) The 0th index of each sequence (the transformed class token) is interpreted as the model prediction. The rest of the sequence is ignored.

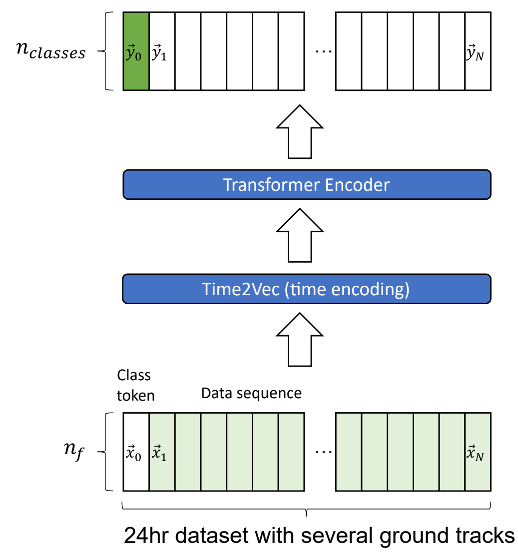

Figure 2: Conceptual diagram of the Measurement Transformer architecture

4 Training Data

The Transformer model presented in this paper is trained using a supervised learning approach. A large database of approximately 20,000 simulated trajectories with finite maneuvers was generated to train the models. Orbit states and spacecraft parameters were randomly chosen from a range of elements:

- Altitude between 400-4,000 km.
- Eccentricity between 0-0.5.
- Inclination between 0-180°.
- All values of longitude of ascending node, argument of perigee, and true anomaly.
- Spacecraft mass between 50-500 kg.

- Drag coefficient between 1.6 and 3.0.
- Finite maneuver thrust between 1-50N.
- Finite maneuver start anytime within a predefined 24-hour window.

For each sample, an initial orbital state and spacecraft properties were randomly sampled from the above uniform distributions then propagated in a high-fidelity force model for 24 hours. Predefined observation platforms recorded simulated measurements whenever the spacecraft is in-view. The list of predefined observers include:

- 3 ground stations collecting coherent (low noise) instantaneous range, Doppler measurements and Azimuth, Elevation optical measurements.
- 5 ground stations collecting noncoherent (high noise) instantaneous range, Doppler measurements and Azimuth, Elevation optical measurements.
- 8 spacecraft in MEO collecting optical Right Ascension, Declination measurements.

Realistic measurement noises are added to the simulated measurements, dependent on the observing platform and observation type.

Table 1: Simulated measurement types and their associated frequencies and noises.

Measurement Type	Frequency	Noise
Ground-based Coherent Doppler	10 s	0.2 mm/s
Ground-based Coherent Range	60 s	1 m
Ground-based Incoherent Doppler	30 s	2 mm/s
Ground-based Incoherent Range	30 s	15 m
Ground-based Az/El	10 s	0.02 deg
Space-based RA/Dec	300 s	0.02 deg

The resultant set of measurements are then processed in a Kalman filter, Monte's UD-factorized sequential batch filter-smoother [3]. To produce a signal in the filter residuals, the estimated maneuver magnitude and direction are purposefully set to incorrect values by choosing an incorrect *a priori* estimate and a small *a prior* variance. Each filter case produces the following, which are used as input features to the Transformer model.

- Epoch
- Measurement type
- Observing station name and state
- Estimated spacecraft state
- Prefit and Postfit residuals

An example of the residuals inputs to the model are plotted in Figure 3.

5 Results

The Transformer model is trained to predict the start time of the maneuver contained in a time series of navigation filter products. To determine whether a maneuver is present within a window of filter products, an anomaly classification model such as those introduced in [5] can first be used. The presented model subsequently seeks to identify at what time in the supplied window the maneuver begins.

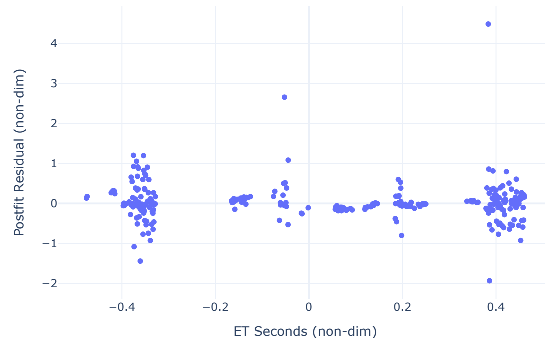

Figure 3: Example of navigation filter postfit residuals containing a maneuver.

For this study, all input sequences are the collection of OD products over the full 24-hour propagation window defined in section 4. However, the proposed architecture is applicable to nearly any realistic window length. The model requires sufficient context before and after the present anomaly, thus the maximum and minimum window lengths are not limited temporally but by data sparsity. Conceptually, the model's performance on short, dense tracking arcs should be comparable to long, sparse arcs, as long as the maneuver is sufficiently represented in each. Since every space mission has different operating conditions, defining absolute bounds on the input time window is not possible.

Furthermore, the proposed model could be deployed to ingest real-time data streams. In this case, the input sequence would be formed from a sliding time window. As navigation data becomes available, new data would be appended to the operating input sequence, the oldest entries would be removed, and the new input sequence would be used for model inference. This process could happen on any time frequency depending on the operator's objectives.

Following machine learning best practices, the model is trained using a 70-20-10, training-validation-test dataset split. During training, the model updates its parameters based on its evluation of the training set. The validation set is passed through the model at the end of each training epoch to ensure the model is not overfitting. The test set is not seen during training and serves to evaluate the performance of the model on new data after training.

5.1 Results for Range and Doppler Measurements

Two-way instantaneous range/Doppler measurements are commonly available for cooperative spacecraft and contain little noise. As an initial study, the Transformer model is trained on navigation products using just range/Doppler measurements to get a baseline understanding of nominal operational performance. The small measurement noises allow the navigation filter to converge more readily and stably, thus any signals present in the filter residuals should be more apparent.

Given a time series of range/Doppler navigation products, the model predicts the non-dimensional maneuver start time within the predefined 24-hour window. The non-dimensional model prediction is transformed back to a dimensional ephemeris time and compared to the true maneuver start. A histogram of the errors between the true maneuver start times and model predictions for the range/Doppler test dataset is shown in Figure 4.

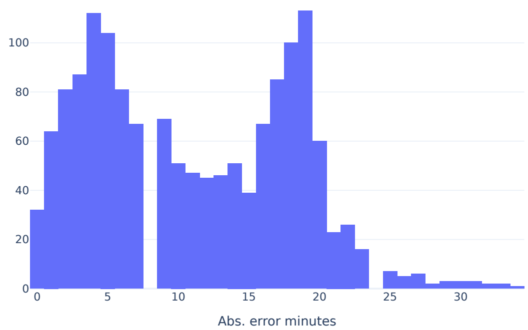

Figure 4: Histogram of the absolute error between model prediction and true start time for range/Doppler products.

Using just range/Doppler measurements, the model is able to accurately predict the start of the maneuver, often to within a few minutes. Of all the test cases, only about 0.5% of the model predictions have an absolute error greater than 30 minutes. Most of the predictions with an error greater than 30 minutes are due to: the measurement occurs in between tracking arcs and/or the tracking is sparse around the start of the maneuver. These cases are visualized in Figure 5 and Figure 6, respectively.

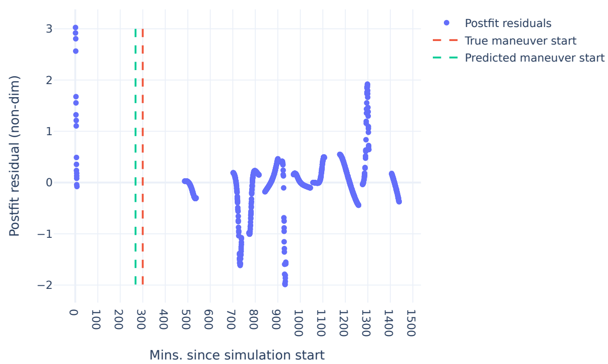

Figure 5: Example of a maneuver occurring in between tracking arcs, producing bad model predictions.

5.2 Results for Ground-based Optical Measurements

High revisit rate range/Doppler measurements are unavailable for uncooperative spacecraft. Several options for tracking these objects exist, such as optical measurements. However, since the spacecraft is uncooperative, the associated measurement noises for all phenomenologies are significant. The larger measurement uncertainties make it more challenging to obtain a converged navigation

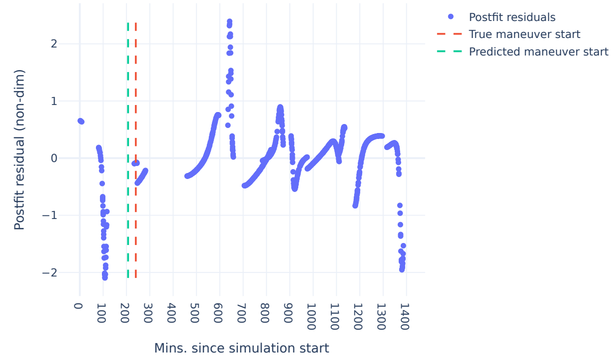

Figure 6: Example of sparse tracking around a maneuver, producing bad model predictions.

solution. Furthermore, signals-of-interest may be lost within the noise of the filter products, especially if the maneuver executed by the spacecraft is small. To investigate the performance of the model for uncooperative maneuver characterization, the model was trained on navigation products from ground-based optical azimuth/elevation measurements. A histogram of the absolute errors between the model predictions and truth values is shown in Figure 7.

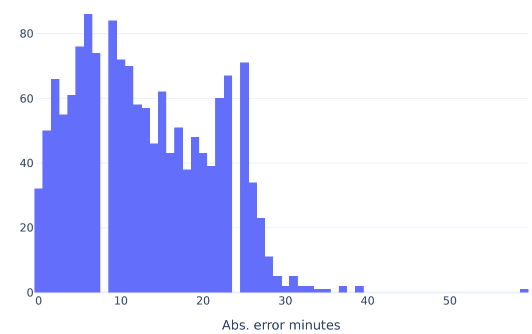

Figure 7: Histogram of the absolute error between model prediction and true start time for ground-based Az/El products.

Despite the lack of cooperative measurements and larger measurement noises, the model can still learn to accurately predict the start of the maneuver to within a few minutes. Approximately 1.0% of the maneuver characterization test cases produce an absolute error of more than 30 minutes.

5.3 Results for Space-based Optical Measurements

Interest in onboard technologies and processing has recently seen a surge in the SSA community. The goal is to eliminate the need to rely on ground stations, computers, and operations. Benefits of onboard SSA technologies include reduced latency, faster decision-making, and improved tracking. To evaluate the NN performance for onboard applications, the model is trained to predict the start of a maneuver in a sequence of filter products generated from space-based right ascension/declination optical measurements. The measurements are simulated between a constellation of spacecraft in MEO collecting observations on a variety of uncooperative spacecraft. The results are shown in Figure 8.

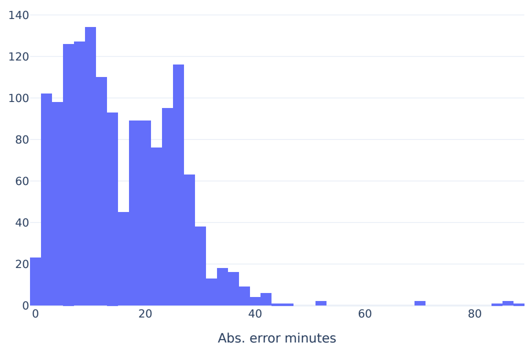

Figure 8: Histogram of the absolute error between model prediction and true start time for space-based RA/DEC products.

When evaluated on simulated space-based optical measurements, the model performance is slightly worse. Approximately 5.8% of the maneuver characterization test cases produce an absolute error of more than 30 minutes. Furthermore, there are more significant outliers with some cases reaching an absolute error of almost 1.5 hours. This reduced performance is most likely due to the lower observation collection frequency of the space-based measurements. While the ground-based optical observations were simulated every 10 seconds, the space-based measurements were only generated once every 300 seconds. As a result, the model has to make a prediction from more sparse tracking data.

6 Conclusion

This study explores the use of an ML model for maneuver characterization from spacecraft navigation data. Specifically, we introduce a Transformer model that, given a 24-hour window of navigation filter products containing a maneuver, predicts the start time of the maneuver. The model is robust to different observer phenomenologies, noncontinuous tracking, and realistic measurement noises. The model is able to determine the start of an unidentified finite burn maneuver to within tens of minutes. Future work will expand the model capabilities to predict more characteristics of the maneuver, such as the maneuver stop time and maneuver magnitude. Coupled with an anomaly detection model, such as those introduced in the authors' prior work [5], the presented model helps enable autonomous SSA operations, such as pattern-of-life analysis.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR* abs/1810.04805 (2018). arXiv:1810.04805 http://arxiv.org/abs/1810.04805

[2] 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. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *CoRR* abs/2010.11929 (2020). arXiv:2010.11929 https://arxiv.org/abs/2010.11929

[3] Scott Evans, William Taber, Theodore Drain, Jonathon Smith, Hsi-Cheng Wu, Michelle Guevara, Richard Sunseri, and James Evans. 2018. MONTE: the next generation of mission design and navigation software. *CEAS Space Journal* 10 (Jan. 2018), 79–86. https://doi.org/10.1007/s12567-017-0171-7

[4] Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus A. Brubaker. 2019. Time2Vec: Learning a Vector Representation of Time. *CoRR* abs/1907.05321 (2019). arXiv:1907.05321 http://arxiv.org/abs/1907.05321

[5] Nathan Parrish Ré, Matthew D. Popplewell, Michael Caudill, Timothy Sullivan, Tyler Hanf, Kanak Parmar, Tyler Presser, Sai Chikine, Michael Grant, and Richard Poulson. 2024. Transformers for Orbit Determination Anomaly Detection and Classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024, AI4SPACE Workshop*. Seattle.

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *CoRR* abs/1706.03762 (2017). arXiv:1706.03762 http://arxiv.org/abs/1706.03762

[7] Qingsong Wen, Tian Zhou, Chaoli Zhang, Weiqi Chen, Ziqing Ma, Junchi Yan, and Liang Sun. 2023. Transformers in Time Series: A Survey. arXiv:2202.07125 [cs.LG] https://arxiv.org/abs/2202.07125