POLITECNICO DI TORINO
Repository ISTITUZIONALE

Physics-Informed Real NVP for Satellite Power System Fault Detection

Original
Physics-Informed Real NVP for Satellite Power System Fault Detection / Cena, C., Albertin, U., Martini, M., Bucci, S., Chiaberge, M.. - ELETTRONICO. - (2024), pp. 679-684. (IEEE International Conference on Advanced Intelligent Mechatronics (AIM) Boston (USA) 15-19 July 2024) [10.1109/AIM55361.2024.10636990].

Availability:
This version is available at: 11583/2991406 since: 2024-08-26T06:44:55Z

Publisher:
IEEE

Published
DOI:10.1109/AIM55361.2024.10636990

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

11 September 2026

arXiv:2405.17339v1 [cs.LG] 27 May 2024

Physics-Informed Real NVP for Satellite Power System Fault Detection

Carlo Cena[1], Umberto Albertin[1], Mauro Martini[1], Silvia Bucci[2] and Marcello Chiaberge[1]

***Abstract*— The unique challenges posed by the space environment, characterized by extreme conditions and limited accessibility, raise the need for robust and reliable techniques to identify and prevent satellite faults. Fault detection methods in the space sector are required to ensure mission success and to protect valuable assets. In this context, this paper proposes an Artificial Intelligence (AI) based fault detection methodology and evaluates its performance on ADAPT (Advanced Diagnostics and Prognostics Testbed), an Electrical Power System (EPS) dataset, crafted in laboratory by NASA.**

Our study focuses on the application of a physics-informed (PI) real-valued non-volume preserving (Real NVP) model for fault detection in space systems. The efficacy of this method is systematically compared against other AI approaches such as Gated Recurrent Unit (GRU) and Autoencoder-based techniques.

Results show that our physics-informed approach outperforms existing methods of fault detection, demonstrating its suitability for addressing the unique challenges of satellite EPS sub-system faults. Furthermore, we unveil the competitive advantage of physics-informed loss in AI models to address specific space needs, namely robustness, reliability, and power constraints, crucial for space exploration and satellite missions.

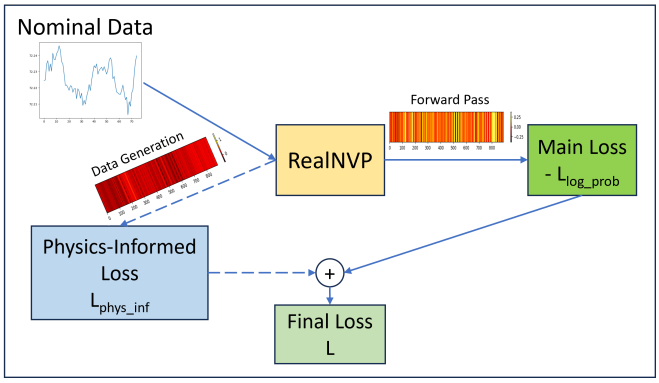

Fig. 1: Training step for the proposed fault detection pipeline with Real NVP. Nominal data is fed into the model and a loss composed of two terms, $-L_{log_prob}$ and L_{phys_inf}, is evaluated. The former, which is the standard loss used with these models, is computed starting from the output of the forward pass. The latter is a physics-informed loss that boosts the performance and interpretability of the model, generated by the inverse propagation of Real NVP.

I. INTRODUCTION

Satellites are essential for many applications such as communication [1], navigation [2], and remote sensing. Earth observation can provide substantial support to monitor climate changes [3], model and predict flood and environmental disasters [4], and extract precious vegetative trends for agriculture [5]. Besides that, space exploration is a rapidly growing economic sector that is pushing the evolution of cutting-edge technology in the satellite field. For example, small size CubeSats have been recently employed for asteroid impact observation [6] and deep space biological exploration [7]. However, operating in the harsh and unpredictable space environment exposes them to various sources of faults that can compromise their functionality and performance [8].

Traditional fault detection methods based on rule-based [9] or model-based [10] techniques have limitations in handling complex and uncertain scenarios, especially when dealing with high-dimensional and nonlinear data.

Therefore, there is a growing interest in applying Artificial Intelligence (AI) methods to enhance the fault detection capabilities of satellites [11], [12]. AI methods can learn from data and extract useful features for fault detection, without relying on prior knowledge or assumptions. Many works tried to exploit time series data patterns to learn fault detection and diagnostic [13], [14], [15], [16], [17]. However, most of the existing AI methods for fault detection in the space sector are based on black-box models that do not incorporate any physical information or constraints. This can lead to unrealistic or inaccurate results, especially when dealing with rare or novel faults. Moreover, black-box models are often computationally expensive and require large amounts of data, which are not always available or feasible in the space domain [18].

In this paper, we propose a novel AI method for fault detection in satellite power systems, based on a physics-informed (PI) [19] real-valued non-volume preserving (Real NVP) model. Real NVP [20] is a type of generative model that can learn the probability distribution of the data and generate realistic samples. By incorporating physical information into the loss function of Real NVP, we can enforce the physical consistency and plausibility of the generated samples, as well as reduce the data and computational requirements. We apply our PI-Real NVP model to the ADAPT dataset [21], an Electrical Power System (EPS) dataset created by NASA in a laboratory setting to simulate various types of faults. We compare the performance of our model with other AI-based methods such as Gated Recurrent

*This publication has received funding from the MUR – DM 351/2022 and is part of the project PNRR-NGEU which has received funding from the MUR – DM 117/2023.

[1]Carlo Cena, Umberto Albertin, Mauro Martini and Marcello Chiaberge are with the Department of Electronics and Telecommunications (DET), Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino (TO), Italy name.surname@polito.it

[2]Silvia Bucci is with Argotec SRL, via Luigi Burgo, 8, 10099 San Mauro Torinese (TO), Italy silvia.bucci@argotecgroup.com

Unit (GRU) [22] and Autoencoder-based techniques [23], [24], and with previous approaches on ADAPT such as the convolutional-based approach proposed in [25] and the Multilayer Perceptron (MLP) adopted in [26]. The fault detection models have been trained with a semi-supervised setup [27]: the training follows the unsupervised paradigm of the Real NVP to learn the nominal data distribution, whilst faulty data are filtered out with human supervision to identify the nominal training data. Faulty data are then used to evaluate the models during testing and validation. We show that our model achieves superior results in terms of accuracy, robustness, and interpretability. We also demonstrate the advantages of using physics-informed loss in AI models for fault detection in space systems, such as reducing the false alarm rate, improving the generalization ability on new types of faults, and saving power consumption. Our work provides a new perspective on how to leverage the synergy between physics and AI for fault detection in satellites and other complex systems.

The contributions of this research are the following:

- We tackle a satellite EPS fault detection problem with a novel semi-supervised Machine Learning approach;
- We introduced a physics-informed loss focused on voltage and current signal relationships of the given EPS sub-system and provided insights into the feature maps produced by the AI models;
- We demonstrate a competitive advantage on the ADAPT dataset outperforming other state-of-the-art methods.

II. Methodology

Here we delve into the training and test phases of each model and describe the losses used, paying particular attention to the physics-informed loss.

We focused on the application of a physics-informed - Real NVP [20] model for fault detection within the ADAPT dataset[1]. In this study, a Real NVP with a multivariate normal distribution is adapted to address the challenges associated with fault detection in the space sector, leveraging its inherent capacity to capture intricate relationships within the data.

Real NVP [20] is a methodology employed in density estimation tasks, particularly within the domain of generative modeling. Developed as an extension of the Normalizing Flow [28], [29] framework, it provides a modeling approach, with tractable likelihood evaluation and efficient sampling, complex high-dimensional distributions by learning an invertible mapping between the data and a simpler latent space.

The key innovation of Real NVP with respect to other Normalising Flow models lies in the use of affine coupling layers (Fig. 2), that can exploit the local dependencies within the data and capture diverse modes present in the data distribution, enabling the generation of realistic and diverse samples.

Each coupling layer operates by decomposing the data into two sets of variables and transforming one set while leaving the other unchanged. This transformation is achieved through a series of invertible, bijective functions, designed such that both forward and inverse mappings are computationally efficient, allowing for tractable and exact inference and generation of samples. These functions are typically implemented as neural networks.

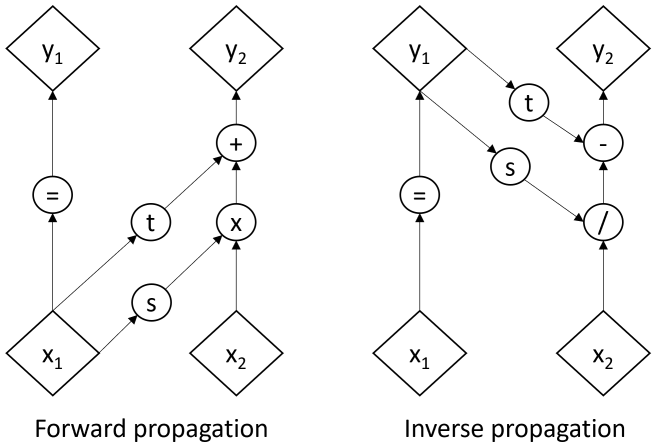

Fig. 2: Computational graph for propagation in both directions in a Coupling Layer [20]. In this kind of model, *s* is related to the scaling function while *t* is related to the translation one. These functions assume the shape of neural networks with several layers that must be designed properly as a function of the application.

Real NVP offers a scalable solution to model multi-modal distributions effectively for density estimation, making it suitable for large-scale datasets.

A. Architectures

Other than the Real NVP model, we implemented a set of other AI models to compare its performances. We analyzed the fault detection literature and selected a couple of the most used approaches: GRU [22] and Autoencoders [23]. GRU was selected as an alternative to the more used Long Short-Term Memory (LSTM) [30], because it has been shown to lead to similar [31] or better [32] performances with respect to the LSTM while being more computationally efficient.

Autoencoder: The Autoencoder [23] used for the test is under-complete. More in detail, the hyper-parameters searched led to an encoder composed of two layers with 256 and 64 neurons, respectively. The decoder comprises two other layers with 256 neurons for the first and as many neurons as the number of input features for the second. Each layer is fully connected with a *relu* activation function. The under-complete Autoencoder reduces the number of input features in a more dense configuration to capture the most salient information regarding the input data.

GRU: [22] The best neural network architecture comprises two GRU layers. Each one has 256 neurons and the head is a fully connected layer with as many neurons as 50 times the input features of the ADAPT dataset, i.e. the time window considered by the model, to reconstruct the input data.

Real NVP: Real NVP [20], whose training procedure is described in the previous section, is a normalizing flow

[1]https://data.nasa.gov/dataset/ADAPT-Dataset/6gjh-n6gb/about_data

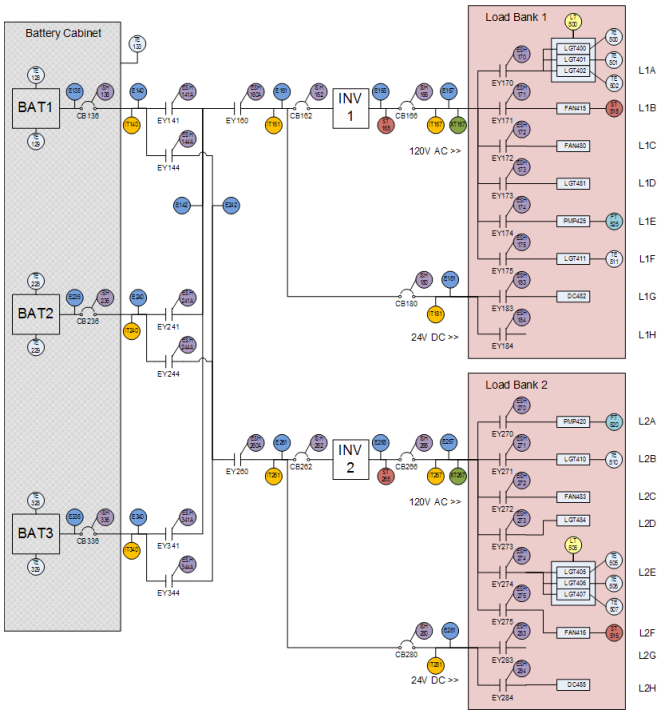

Fig. 3: Circuit of the testbed used to generate the ADAPT dataset. This circuit is used to handle the constraints related to the physics-informed loss composition.

model. In this model, the neural network to be designed regards the scale *s* and translation *t* functions. *t* is designed with four fully connected layers with 128 neurons and a fully connected layer with a number of neurons equal to the input features. The *s* function is designed with the same structure as the *t* transformation with a *tanh* activation function for the last layer instead of the *linear* activation function of the *t*.

B. Loss Functions

Moreover, we designed a physics-informed loss, to be used with all models, extracting physical relations from the circuit of the testbed assembled by NASA to create ADAPT, shown in Fig. 3. This loss is composed of the following terms:

$$d_{squared} = \frac{1}{|\boldsymbol{O}|} \sum_{i \in |\boldsymbol{O}|} \begin{cases} (x_i - 1)^2, & \text{if } x_i > 1 \\ x_i^2, & \text{if } x_i < 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$e = \frac{\sum_{ij} \overline{(\boldsymbol{E_i} - \boldsymbol{E_j})^2}}{M} \quad (2)$$

$$it = \frac{\sum_i \overline{\boldsymbol{IT_i}^2}}{N} \quad (3)$$

$$E65 = \frac{\overline{|\boldsymbol{E165} - \boldsymbol{v}|} + \overline{|\boldsymbol{E265} - \boldsymbol{v}|}}{2} \quad (4)$$

$$ST65 = \frac{\overline{|\boldsymbol{ST165} - \boldsymbol{f}|} + \overline{|\boldsymbol{ST265} - \boldsymbol{f}|}}{2} \quad (5)$$

$d_{squared}$ is used to steer all the values of the array $\boldsymbol{O}$, generated through the Real NVP inverse propagation, between 0 and 1. e has the purpose of forcing M pairs of voltages, $\boldsymbol{E_i}$ and $\boldsymbol{E_j}$, to be equal, with M given by the number of pairs with a closed circuit breaker/relay between them (An example is the outputs of voltage sensors $\boldsymbol{E_i} = \boldsymbol{E135}$ and $\boldsymbol{E_j} = \boldsymbol{E140}$, separated by the circuit breaker CB136 in the high-left corner of Fig. 3), *it* is designed to set to 0 all N currents, $\boldsymbol{IT_i}$, measured in an open circuit, whereas $E65$ and $ST65$ steer a couple of voltages and frequencies to the values of 120.5V ($\boldsymbol{v}$) and 60Hz ($\boldsymbol{f}$), respectively, given their proximity to INV1 and INV2.

The vectors $\boldsymbol{E}$, $\boldsymbol{IT}$, $\boldsymbol{E165}$, $\boldsymbol{E265}$, $\boldsymbol{ST165}$, $\boldsymbol{ST265}$, $\boldsymbol{v}$, and $\boldsymbol{f}$ have a length equal to the time window given in input to the models, one of the hyper-parameters tuned, as described in the following section.

The physics-informed loss is equal to

$$L_{phys_inf} = d_{squared} + e + it + E65 + ST65 \quad (6)$$

The total loss is equal to:

$$L = L_{main} + \beta \cdot L_{phys_inf} \quad (7)$$

Where β is a weight learned through the Lagrangian dual framework [33], which uses a Lagrangian dual approach to learn the best multipliers of a Lagrangian loss with a sub-gradient method.

L_{main} is the main loss for each model, in the case of the GRU and Autoencoder models respectively the Mean Absolute Error (MAE) and the Mean Squared Error (MSE) were used, while for the Real NVP models $L_{main} = -L_{log_prob}$, where L_{log_prob} is the log-likelihood of the current data.

The physics-informed loss was applied on the reconstructed array produced in output by the GRU and the Autoencoder, while, given that the Real NVP can be used to generate new data starting from a probability distribution, we applied this loss on the generated arrays. In each training step the Real NVP models can generate an arbitrary number of arrays, this number was selected through hyper-parameter tuning.

C. Training & Test

In Fig. 1 and 4 we show how the models were trained. This was done in two different ways: all models received in input the nominal data, but while the GRU and the Autoencoder were trained to minimize the reconstruction error, the Real NVP was trained to increase the log-likelihood of the input data coming from the output distribution.

In Fig. 5 is shown a graphical representation of how the above-mentioned Physics-Informed loss works.

During inference the output of the GRU and Autoencoder models was used to compute a reconstruction error, which can then be compared to a threshold to separate nominal and fault data, whereas the output of the Real NVP models, being a log-likelihood, can be directly used to separate nominal data from fault data, using a threshold.

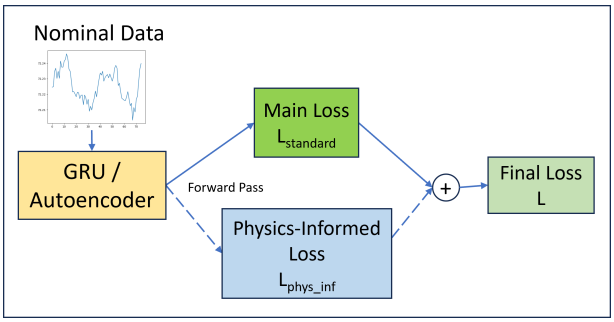

Fig. 4: Training method for GRU and Autoencoders. The models are tested with standard and physics-informed loss.

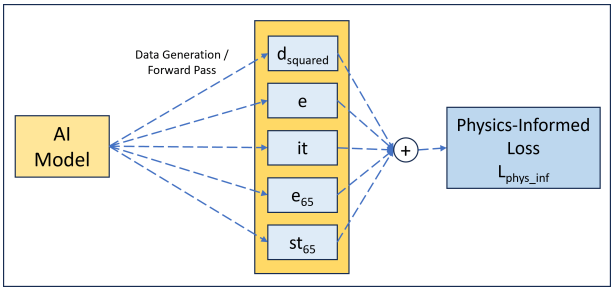

Fig. 5: Physics-informed loss composition (6).

III. Experiments & Results

In the following, we first present how we conducted the experiments. Then, we report and analyze the results obtained.

A. Experimental Setup

Here we describe the dataset, the metrics used to evaluate the effectiveness of the proposed solution for the specific task, the libraries, and the computational resources used to perform the experiments.

Dataset: The dataset used for testing the model is ADAPT [21], provided by NASA Ames Research Center. This dataset has been created with an EPS testbed to evaluate fault detection and isolation algorithms using controlled and repeatable fault injection scenarios. The dataset contains healthy and faulty conditions representative of the EPSs used in the aerospace sector.

Metrics: To evaluate the effectiveness of the Autoencoder, GRU, and Real NVP in detecting fault data in both dataset's splits and ensure their robustness and applicability, we consider the Area Under the ROC Curve (AUROC), from now on also abbreviated as AUC, False Positive Rate at 95% true positive rate (FPR95), and F1-score.

AUC gives a comprehensive performance summary and is robust to imbalanced data, FPR95 gives a measure of the false positive rate at high specificity (95%), this is particularly important in applications where false positives can have high costs. Finally, the F1-score is a typically used metric added here to ease comparisons. AUC and FPR95 let us analyze the performance of our models without selecting a threshold. While we use the threshold at FPR95 to compute the F1-score.

Model	Past Length	Coup. Layers	Layers	Neurons	Batch Size
GRU - AE	50, 30, 10	-	2,4,6,8	512,256, 128,64,32	256,128, 64,32
Real NVP	50, 30, 10	2,4,6	2,4,6,8	512,256, 128,64,32	256,128, 64,32

TABLE I: Grid search combinations tested for each model. The coupling layers are used only for the Real NVP model.

Implementation Details: We started with a comprehensive pre-processing phase, where the dataset was cleaned, scaled between 0 and 1, and two train-test splits were generated as described below.

A set of 10 random 70-30% splits were created and a coarse grid search was launched using each one of them. Each grid search was performed using 30% of the training data as validation set. Finally, the two more challenging splits that led to the worst AUC results on the test set were selected and used in all subsequent steps. This was done because an initial exploratory analysis and previous works [26], [25] showed that it was quite easy to obtain good results on ADAPT.

The training of all models comprised an initial phase of coarse grid search and a second phase of fine-grained grid search with 5 different random seeds.

During these phases, we optimized different hyper-parameters based on the model at hand, including the amount of past data used as input (i.e. the length of the input array), the number of layers, either fully-connected or GRU, and neurons of the NNs, the batch size, the number of coupling layers, and whether or not to use the physics-informed loss. All the ranges used for the hyper-parameters tuning are shown in Table I.

To perform these experiments we used custom models created with TensorFlow[2]. Each experiment was run on an Intel Xeon Scalable Processors Gold 6130 with 16 cores and a frequency of 2.10 GHz.

B. Results

Table II summarizes the results obtained for both the dataset splits with the above-mentioned methods by averaging 5 different seeds during the grid search hyper-parameters optimization. The results reported in the table show the applications of both physics and not physics-informed loss. The PI-Real NVP works better in both splits. As shown in the table, the AUC of the ROC curve obtained from the PI-Real NVP is similar to the one of the Autoencoder; on the other hand, the FPR95 is lower, and the F1-score is higher in Real NVP. FPR95 is a critical score because it denotes how many false alarms the maintainer receives from the system. These scores confirm a reduction in the false positive detection with respect to other methods, making the PI-Real NVP more robust.

Moreover, it can be noted that the models trained with the physics-informed loss achieve better performances with

[2]https://www.tensorflow.org/

Split	Model	Past Length	AUC	FPR95	F1-score
1	Real NVP	50	0.94	0.142	0.95
	PI-Real NVP	50	**0.95**	**0.101**	**0.95**
	GRU	50	0.91	0.301	0.92
	PI GRU	50	0.93	0.249	0.93
	Autoencoder	50	0.94	0.142	0.94
	PI Autoencoder	50	0.95	0.108	0.95
	kPCA + MLP [26]	-	0.93	0.504	0.74
2	Real NVP	50	0.99	0.047	0.96
	PI-Real NVP	**10**	**0.99**	**0.005**	**0.98**
	GRU	50	0.96	0.146	0.94
	PI GRU	50	0.96	0.135	0.94
	Autoencoder	50	0.99	0.063	0.95
	PI Autoencoder	50	0.99	0.041	0.96
	kPCA + MLP [26]	-	0.93	0.111	0.92

TABLE II: Results obtained for every split with each model averaged over 5 random seeds. For each result, we also show the amount of past data used. The PI-Real NVP model outperforms all the other models. Our model shows the best results in the second split, using less past data.

Split	Sensor	Model	Past Length	AUC	FPR95	F1-score
1	ST515	PI-Real NVP	**50**	**0.92**	**0.08**	**0.81**
	ST515	ST + CNN [25]	120	0.77	0.93	0.15
2	E161	PI-Real NVP	**10**	**0.99**	**0.00**	**0.98**
	E161	ST + CNN [25]	120	0.41	0.66	0.18
	ST515	PI-Real NVP	10	0.99	0.00	0.98
	ST515	ST + CNN [25]	120	0.99	0.00	0.98

TABLE III: Results obtained when applying the method of [25] on the columns of our splits with faulty data both in the train and test data. We tested our best models by removing from the test set all faults but those of the considered sensor. PI-Real NVP achieves superior or equivalent results.

respect to the baselines, with the exception of the GRU models in the second split, which led to a slight improvement only for the FPR95. This shows that the physics-informed loss aids the fault detection process: it leads to a higher percentage of detected faults and reduces the number of false positives, thus providing more thrusthworty AI solutions.

Compared to previous AI studies on ADAPT, our models (and loss) achieve better performances with respect to the multivariate supervised approach [26], as shown in Table II, and to the univariate supervised approach [25] reported in Table III.

[25] uses a CNN, preceded by a Stockwell Transform, to perform fault detection on a single sensor, i.e. a single column. It trains and tests each model only on a column of the dataset, thus dealing with an univariate distribution. While using a classifier for each sensor simplifies the data distribution and automatically covers also the fault isolation step, it requires to run inference on significantly more classifiers, thus leading to higher memory and computational costs. To compare our results to those of [25], for each split we: (1) selected the sensors with faulty measurements both in the train and in the test data; (2) trained and tested their method on each of those sensors data, performing a grid search for the parameters that we couldn't find in [25], i.e. kernel size $\{1, 3, 5, 7\}$ and frequency index range $\{100, 300, 500\}$; (3) compared the results obtained with those of our best model (PI-Real NVP) evaluated by removing from the test set all anomalies except those linked to the considered sensor.

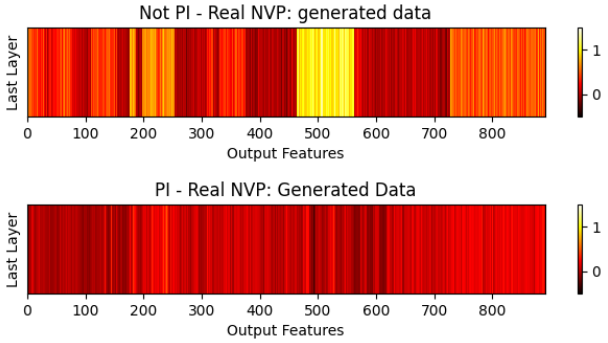

Fig. 6: Data generated by the inverse propagation of the not PI-Real NVP, at the top, and the PI-Real NVP, at the bottom. On the x-axis are shown the output features for an experiment with a past window length of 10. These figures show that the PI loss successfully steered the Real NVP model towards the generation of data that doesn't violate the physics of the circuit, i.e. which falls in the range 0-1.

[26] uses a MLP trained in a supervised fashion after having pre-processed the dataset with Kernel Principal Component Analysis (KPCA). Supervised approaches cannot be efficiently extended to datasets without, or with a very limited amount of labels as is often the case for satellite applications. In turn, our methods, and in particular Real NVP, which have been trained in a semi-supervised fashion, can be easily extended to incorporate a supervised loss. To train it we performed a grid search for the number of units of the fully-connected layers $\{512, 256, 128, 64\}$ and for the batch size $\{64, 32, 16\}$, because we weren't able to find them in [26].

Feature Maps Insights: To provide a few insights on the effects of the physics-informed loss, we show the output of the last layer for the best performing model, i.e. the coupling layers of Real NVP, with and without PI loss. Fig. 6 shows the feature maps generated by the inverse propagation, i.e. when generating new data, of the non-PI and the PI models, respectively at the top and at the bottom of the figure. It can be noted that the features generated by the model trained without PI loss violate the physics of the circuit, as they are expected to be in the range 0-1, while those generated by the PI model fall in the given range. This demonstrates that the Real NVP successfully learned the $d_{squared}$ sub-reward (1) and shows that the use of PI losses provides the means to improve the performance and explainability of the given AI model, resulting in a subsequent boost in confidence towards it.

IV. Conclusions & Future Works

This paper presents an application of Physics-Informed - Real NVP for Fault Detection. As explained in the previous sections, the experiments are performed on the ADAPT

dataset by designing a physic-informed loss, which steers the model learning by providing physical insights into the data. This study shows that using PI-Real NVP leads to improved results on an EPS dataset built to be representative of satellite data coming from the same sub-system. Additionally, we showed that models trained with our physics-informed loss perform better with respect to those trained only with a standard loss. Given the high performances obtained in this dataset with a large set of AI models, we plan on testing the PI-Real NVP approach further by performing ablation studies on several parameters belonging both to the dataset and the models (e.g. number of input features, number of training data), and subsequently test it on new datasets representative both of EPS and of other satellites' sub-systems.

ACKNOWLEDGMENT

This work has been developed with the contribution of the Politecnico di Torino Interdepartmental Centre for Service Robotics (PIC4SeR https://pic4ser.polito.it) and Argotec SRL, through the MUR – DM 117/2023 scholarship of Carlo Cena. Computational resources were provided by HPC@POLITO (http://hpc.polito.it).

REFERENCES

[1] Y. Rahmat-Samii and A. C. Densmore, "Technology trends and challenges of antennas for satellite communication systems," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 4, pp. 1191–1204, 2014.

[2] G. W. Hein, "Status, perspectives and trends of satellite navigation," *Satellite Navigation*, vol. 1, no. 1, p. 22, 2020.

[3] D. Ballari, L. M. Vilches-Blázquez, M. L. Orellana-Samaniego, F. Salgado-Castillo, A. E. Ochoa-Sánchez, V. Graw, N. Turini, and J. Bendix, "Satellite Earth Observation for Essential Climate Variables Supporting Sustainable Development Goals: A Review on Applications," *Remote Sensing*, vol. 15, no. 11, 2023.

[4] G. Schumann, L. Giustarini, A. Tarpanelli, B. Jarihani, and S. Martinis, "Flood Modeling and Prediction Using Earth Observation Data," *Surveys in Geophysics*, vol. 44, no. 5, pp. 1553–1578, Oct 2023.

[5] M. Martini, V. Mazzia, A. Khaliq, and M. Chiaberge, "Domain-adversarial training of self-attention-based networks for land cover classification using multi-temporal sentinel-2 satellite imagery," *Remote Sensing*, vol. 13, no. 13, p. 2564, 2021.

[6] E. Dotto and A. Zinzi, "Impact observations of asteroid Dimorphos via Light Italian CubeSat for imaging of asteroids LICIACube," *Nature Communications*, vol. 14, no. 1, p. 3055, May 2023.

[7] S. Massaro Tieze, L. C. Liddell, S. R. Santa Maria, and S. Bhattacharya, "BioSentinel: A Biological CubeSat for Deep Space Exploration," *Astrobiology*, vol. 23, no. 6, pp. 631–636, Apr. 2020.

[8] Y. Xu, X. Kong, W. Ma, Z. Zhu, and J. Zhang, "Integrated System Health Management Research and Application on Satellite," in *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)*, 2021, pp. 1–5.

[9] J. Schein, S. T. Bushby, N. S. Castro, and J. M. House, "A rule-based fault detection method for air handling units," *Energy and buildings*, vol. 38, no. 12, pp. 1485–1492, 2006.

[10] R. Isermann, "Model-based fault-detection and diagnosis–status and applications," *Annual Reviews in control*, vol. 29, no. 1, pp. 71–85, 2005.

[11] J. Yu and Y. Zhang, "Challenges and opportunities of deep learning-based process fault detection and diagnosis: a review," *Neural Computing and Applications*, vol. 35, no. 1, pp. 211–252, Jan 2023.

[12] M. Bieber, W. J. C. Verhagen, F. Cosson, and B. F. Santos, "Generic Diagnostic Framework for Anomaly Detection–Application in Satellite and Spacecraft Systems," *Aerospace*, vol. 10, no. 8, 2023.

[13] M. I. Mohamed, K. M. Badran, and A. E. Hussien, "Anomaly Detection for Agile Satellite Attitude Control System Using Hybrid Deep-Learning Technique," *Journal of Aerospace Information Systems*, vol. 20, no. 12, pp. 890–904, 2023.

[14] B. Yu, Y. Yu, J. Xu, G. Xiang, and Z. Yang, "MAG: A Novel Approach for Effective Anomaly Detection in Spacecraft Telemetry Data," *IEEE Transactions on Industrial Informatics*, pp. 1–9, 2023.

[15] S. Feng, C. Miao, K. Xu, J. Wu, P. Wu, Y. Zhang, and P. Zhao, "Multi-scale attention flow for probabilistic time series forecasting," *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[16] W. Sun, A. R. Paiva, P. Xu, A. Sundaram, and R. D. Braatz, "Fault detection and identification using Bayesian recurrent neural networks," *Computers & Chemical Engineering*, vol. 141, p. 106991, 2020.

[17] S. Chen, J. G., and X. Ma, "Satellite on-orbit anomaly detection method based on a dynamic threshold and causality pruning," *IEEE Access*, vol. 9, pp. 86 751–86 758, 2021.

[18] T. Aziz, G. Salama, M. Mohamed, and S. Hussein, "Spacecraft fault detection and identification techniques using artificial intelligence," *Journal of Physics: Conference Series*, vol. 2616, p. 012025, 11 2023.

[19] S. A. Faroughi, N. M. Pawar, C. Fernandes, M. Raissi, S. Das, N. K. Kalantari, and S. Kourosh Mahjour, "Physics-Guided, Physics-Informed, and Physics-Encoded Neural Networks and Operators in Scientific Computing: Fluid and Solid Mechanics," *Journal of Computing and Information Science in Engineering*, vol. 24, no. 4, p. 040802, 01 2024.

[20] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using Real NVP," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[21] O. J. Mengshoel, A. Darwiche, K. Cascio, M. Chavira, S. Poll, and N. S. Uckun, "Diagnosing Faults in Electrical Power Systems of Spacecraft and Aircraft," in *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 1699–1705.

[22] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734.

[23] D. E. Rumelhart and J. L. McClelland, *Learning Internal Representations by Error Propagation*. MIT Press, 1987, pp. 318–362.

[24] M. Ganesan and R. Lavanya, "Simultaneous fault detection in satellite power systems using deep autoencoders and classifier chain," *Telecommunication Systems*, vol. 83, no. 1, pp. 1–15, May 2023.

[25] M. Ganesan, R. V. Lavanya, and M. N. Devi, "Fault detection in satellite power system using convolutional neural network," *Telecommunication Systems*, vol. 76, pp. 505 – 511, 2020.

[26] S. Dheepadharshani, S. Anandh, K. B. Bhavinaya, and R. Lavanya, "Multivariate Time-series Classification for Automated Fault Detection in Satellite Power Systems," in *2019 International Conference on Communication and Signal Processing (ICCSP)*, 2019, pp. 0814–0817.

[27] M. Goldstein and S. Uchida, "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data," *PLOS ONE*, vol. 11, no. 4, pp. 1–31, 04 2016.

[28] E. G. Tabak and E. Vanden-Eijnden, "Density estimation by dual ascent of the log-likelihood," *Communications in Mathematical Sciences*, vol. 8, no. 1, pp. 217 – 233, 2010.

[29] E. G. Tabak and C. V. Turner, "A Family of Nonparametric Density Estimation Algorithms," *Communications on Pure and Applied Mathematics*, vol. 66, no. 2, pp. 145–164, 2013.

[30] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, p. 1735–1780, nov 1997.

[31] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[32] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, p. 2342–2350.

[33] F. Fioretto, P. Van Hentenryck, T. W. K. Mak, C. Tran, F. Baldo, and M. Lombardi, "Lagrangian Duality for Constrained Deep Learning," in *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track*, Y. Dong, G. Ifrim, D. Mladenić, C. Saunders, and S. Van Hoecke, Eds. Cham: Springer International Publishing, 2021, pp. 118–135.