ESA Anomaly Detection Benchmark
ESA 异常检测基准

The European Space Agency Anomaly Detection Benchmark (ESA-ADB) consists of three main components (visualised in the figure below for easier comprehension):

欧洲航天局异常检测基准 （ESA-ADB） 由三个主要组成部分组成（如下图所示，以便于理解）：

Large-scale, curated, structured, ML-ready ESA Anomalies Dataset (ESA-AD, in short) of real-life satellite telemetry collected from three ESA missions (out of which two are selected for benchmarking in ESA-ADB), manually annotated by spacecraft operations engineers (SOEs) and ML experts, and cross-verified using state-of-the-art algorithms. It can be downloaded from here: https://doi.org/10.5281/zenodo.12528696

从三个 ESA 任务（其中两个被选中用于 ESA-ADB 的基准测试）收集的大规模、精心策划、结构化、ML 就绪的 ESA 异常数据集（简称 ESA-AD），由航天器运营工程师 （SOE） 和 ML 专家手动注释，并使用最先进的算法进行交叉验证。可以从这里下载： https://doi.org/10.5281/zenodo.12528696

Evaluation pipeline designed by ML experts for the practical needs of SOEs from the ESA's European Space Operations Centre (ESOC). It introduces new metrics designed for satellite telemetry according to the latest advancements in time series anomaly detection (TSAD) and simulates real operational scenarios, e.g. different mission phases and real-time monitoring.

由欧空局欧洲空间运营中心 （ESOC） 的 ML 专家为满足国有企业的实际需求设计的评估管道。它根据时间序列异常检测 （TSAD） 的最新进展引入了为卫星遥测设计的新指标，并模拟了真实的作场景，例如不同的任务阶段和实时监控。

Benchmarking results of TSAD algorithms selected and improved to comply with the space operations requirements.

选择和改进的 TSAD 算法的基准测试结果，以符合空间作要求。

We hope that this unique benchmark will allow researchers and scientists from academia, research institutes, national and international space agencies, and industry to validate models and approaches on a common baseline as well as research and develop novel, computational-efficient approaches for anomaly detection in satellite telemetry data.

我们希望这一独特的基准将使来自学术界、研究机构、国家和国际航天机构以及工业界的研究人员和科学家能够在共同基线上验证模型和方法，并研究和开发用于卫星遥测数据异常检测的新型、计算高效的方法。

The dataset results from the work of an 18-month project carried by an industry Consortium composed of Airbus Defence and Space, KP Labs, and the European Space Agency's European Space Operations Centre. The project, funded by the European Space Agency (ESA) under the contract number 4000137682/22/D/SR, is a part of The Artificial Intelligence for Automation (A^{2}I) Roadmap (De Canio et al., Development of an actionable AI roadmap for automating mission operations, 2023 SpaceOps Conference), a large endeavour started in 2021 to

automate space operations by leveraging artificial intelligence.

该数据集是由空中客车防务与航天公司、KP 实验室和欧洲航天局欧洲太空运营中心组成的行业联盟开展的为期 18 个月的项目工作的结果。该项目由欧洲航天局 （ESA） 资助，合同编号为 4000137682/22/D/SR，是自动化人工智能 （A²I） 路线图的一部分（De Canio 等人，开发用于自动化任务运营的可作人工智能路线图，2023 年太空行动会议），一项于 2021 年开始的一项大型努力，旨在利用人工智能实现太空运营自动化。

The introduction below describes how to reproduce results presented in the ESA-ADB paper using the provided modified fork of the TimeEval framework.

下面的介绍了如何使用提供的 TimeEval 框架的修改分支重现 ESA-ADB 论文中呈现的结果。

Initial requirements
初始要求

- Some ESA-ADB functions work only on Linux (or Windows Subsystem for Linux 2), so we suggest using it as a development platform

  某些 ESA-ADB 函数仅适用于 Linux（或适用于 Linux 2 的 Windows 子系统），因此我们建议将其用作开发平台
- The hard disk must be configured to use NTFS file system

  硬盘必须配置为使用 NTFS 文件系统
- It is recommended to have at least 512 GB free disk space to store artifacts from all experiments

  建议至少有 512 GB 的可用磁盘空间来存储所有实验的工件
- It is recommended to use Nvidia GPU with compute capability >= 7.1

  建议使用具有计算能力 >= 7.1 的 Nvidia GPU
- It is recommended to use a machine with at least 64 GB RAM (32 GB is an absolute minimum for the whole pipeline)

  建议使用至少具有 64 GB RAM 的机器（32 GB 是整个管道的绝对最低值）

For Windows it is recommended to run the following command before cloning to prevent issues when running the docker containers: `git config --global core.autocrlf false`.

对于 Windows，建议在克隆之前运行以下命令，以防止在运行 docker 容器时出现问题： `git config --global core.autocrlf false` 。

Environment setup
环境设置

Python environment

Python 环境

Install Anaconda environment manager, version >= 22.

安装 Anaconda 环境管理器，版本 >= 22。

Create a conda-environment and install all required dependencies. Use the file `environment.yml` for this: `conda env create --file environment.yml`. Note that you should **not** install TimeEval from PyPI. Our repository contains the modified version of TimeEval in the "timeeval" folder.

创建一个 conda-environment 并安装所有必需的依赖项。为此使用文件 `environment.yml`： `conda env create --file environment.yml` 。请注意，您**不**应该从 PyPI 安装 TimeEval。我们的存储库在“timeeval”文件夹中包含 TimeEval 的修改版本。

Activate the new environment `conda activate timeeval`.

激活新环境 `conda activate timeeval`。

(This step is optional and should be used only if you face any problems with "import timeeval" in your environment and you do not plan to modify the code of TimeEval. Otherwise, the recommended way is to just run scripts from the main repo folder, so our timeeval folder is directly visible to them. Another option is to add the main repo folder to the system PATH) Install the local version of TimeEval: `python setup.py install`

（此步骤是可选的，仅当您在环境中遇到“import timeeval”的任何问题并且不打算修改 TimeEval 的代码时才应使用。否则，推荐的方法是只从主存储库文件夹运行脚本，这样我们的 timeeval 文件夹对他们来说是直接可见的。另一种选择是将主存储库文件夹添加到系统 PATH）安装本地版本的 TimeEval： `python setup.py install`

Docker

码头工人

Install Docker Engine, version >= 23.

安装 Docker 引擎，版本 >= 23。

Build Docker containers with algorithms of interest (e.g., listed in mission1_experiments.py) using instruction from README in the TimeEval-algorithms folder. For our Telemanom-ESA, it is enough to run `sudo docker build -t registry.gitlab.hpi.de/akita/i/telemanom_esa ./telemanom_esa`. For our DC-VAE-ESA, it is enough to run `sudo docker build -t registry.gitlab.hpi.de/akita/i/dc_vae ./dc_vae`.

使用 TimeEval-algorithms 文件夹中的 README 中的指令，使用感兴趣的算法（例如，mission1_experiments.py 中列出）构建 Docker 容器。对于我们的 Telemanom-ESA 来说，运行 `sudo docker build -t registry.gitlab.hpi.de/akita/i/telemanom_esa ./telemanom_esa` .对于我们的 DC-VAE-ESA 来说，运行 `sudo docker build -t registry.gitlab.hpi.de/akita/i/dc_vae ./dc_vae` .

Preparing datasets
准备数据集

Download raw ESA Anomalies Dataset from the link https://doi.org/10.5281/zenodo.12528696 and put ESA-Mission1 and ESA-Mission2 folders in the "data" folder.

从链接 https://doi.org/10.5281/zenodo.12528696 下载原始 ESA 异常数据集，并将 ESA-Mission1 和 ESA-Mission2 文件夹放在“data”文件夹中。

Generating preprocessed data for experiments
生成用于实验的预处理数据

There are separate script to generate preprocessed data for TimeEval framework for each mission. The scripts are located in notebooks\data-prep folder. From the notebooks\data-prep folder run:

有单独的脚本可以为每个任务的 TimeEval 框架生成预处理数据。脚本位于 notebooks\data-prep 文件夹中。在 notebooks\data-prep 文件夹中运行：

Mission1:

任务1：

```
python Mission1_semisupervised_prep_from_raw.py ../../data/ESA-Mission1
```

Mission2:

任务2：

```
python Mission2_semiunsupervised_prep_from_raw.py ../../data/ESA-Mission2
```

The scripts generate all necessary files to data/preprocessed/multivariate folders and add records to data/preprocessed/datasets.csv if necessary (records for ESA-ADB are already added as a part of this repository). Note that the preprocessing may take a few hours on a standard PC.

脚本将所有必要的文件生成到 data/preprocessed/multivariate 文件夹，并在必要时将记录添加到 data/preprocessed/datasets.csv（ESA-ADB 的记录已作为此存储库的一部分添加）。请注意，在标准 PC 上，预处理可能需要几个小时。

Running experiments
运行实验

There is a separate script in the main folder of the repo to run a full grid of experiments for each mission:

存储库的主文件夹中有一个单独的脚本，用于为每个任务运行完整的实验网格：

- Mission1: mission1_experiments.py

  任务 1：mission1_experiments.py

- Mission2: mission2_experiments.py

  任务 2：mission2_experiments.py

The scripts configure and run all algorithms in Docker containers. Results are generated to 'results' folder. On a standard PC, it may be necessary to run algorithms separately, one by one.

这些脚本在 Docker 容器中配置和运行所有算法。结果生成到“results”文件夹中。在标准 PC 上，可能需要逐个单独运行算法。

Notes

笔记

- evaluation pipeline with novel time-aware metrics can only be run for datasets following the same structure as ESA Anomalies Dataset (with labels.csv and anomaly_types.csv)

  具有新时间感知指标的评估管道只能对遵循与 ESA 异常数据集相同结构的数据集运行（具有 labels.csv 和 anomaly_types.csv）

- when analyzing results for different anomaly types for the lightweight subsets of channels, it is necessary to regenerate anomaly types using anomaly_types.csv using scripts/infer_anomaly_types.py. It is because anomaly types may depend on the analyzed subset of channels.

  在分析通道轻量级子集的不同异常类型的结果时，有必要使用脚本/infer_anomaly_types.py 使用 anomaly_types.csv 重新生成异常类型。这是因为异常类型可能取决于分析的通道子集。

- for now, all algorithms treat rare nominal events as anomalies. To change that behaviour, it would be necessary to modify the code of the framework and some algorithms

  目前，所有算法都将罕见的名义事件视为异常。要改变这种行为，有必要修改框架的代码和一些算法

TimeEval

时间评估

The code of the benchmark is based on the TimeEval framework. Please refer to its documentation in case of any detailed questions about API.

基准测试的代码基于 TimeEval 框架 。如果有任何有关 API 的详细问题，请参阅其文档。

TimeEval Citation
TimeEval 引文

If you use TimeEval in your project or research, please cite the demonstration paper:

如果您在项目或研究中使用 TimeEval，请引用演示论文：

> Phillip Wenig, Sebastian Schmidl, and Thorsten Papenbrock. TimeEval: A Benchmarking Toolkit for Time Series Anomaly Detection Algorithms. PVLDB, 15(12): 3678 - 3681, 2022. doi:10.14778/3554821.3554873
>
> 菲利普·韦尼格、塞巴斯蒂安·施密德尔和托尔斯滕·帕彭布洛克。TimeEval：时间序列异常检测算法的基准测试工具包。PVLDB， 15（12）： 3678 - 3681， 2022.土井：10.14778/3554821.3554873

```
@article{WenigEtAl2022TimeEval,
  title = {TimeEval: {{A}} Benchmarking Toolkit for Time Series Anomaly Detection Algorithms},
  author = {Wenig, Phillip and Schmidl, Sebastian and Papenbrock, Thorsten},
  date = {2022},
  journaltitle = {Proceedings of the {{VLDB Endowment}} ({{PVLDB}})},
  volume = {15},
  number = {12},
  pages = {3678--3681},
  doi = {10.14778/3554821.3554873}
}
```

ESA-ADB Citation
欧空局-亚行引文

If you refer to ESA-ADB in your work, please cite our paper:

如果您在工作中引用 ESA-ADB，请引用我们的论文：

> Krzysztof Kotowski, Christoph Haskamp, Jacek Andrzejewski, Bogdan Ruszczak, Jakub Nalepa, Daniel Lakey, Peter Collins, Aybike Kolmas, Mauro Bartesaghi, Jose Martínez-Heras, and Gabriele De Canio. European Space Agency Benchmark for Anomaly Detection in Satellite Telemetry. arXiv, 2024. doi:10.48550/arXiv.2406.17826
>
> 克日什托夫·科托夫斯基、克里斯托夫·哈斯坎普、雅切克·安杰耶夫斯基、博格丹·鲁什恰克、雅库布·纳莱帕、丹尼尔·莱基、彼得·柯林斯、艾比克·科尔马斯、毛罗·巴尔特萨吉、何塞·马丁内斯-赫拉斯和加布里埃尔·德卡尼奥。欧洲航天局卫星遥测异常检测基准。arXiv，2024 年。doi：10.48550/arXiv.2406.17826

```
@article{kotowski_european_2024,
  title = {European {Space} {Agency} {Benchmark} for {Anomaly} {Detection} in {Satellite} {Telemetry}},
  author = {Kotowski, Krzysztof and Haskamp, Christoph and Andrzejewski, Jacek and Ruszczak, Bogdan and
  date = {2024},
  publisher = {arXiv},
```

```
  doi = {10.48550/arXiv.2406.17826}
}
```