SpaceOps-2023, ID #160

An Artificial Intelligence Framework for Space Operations Data and Model Management

Rohaan Ahmed[a]*, Andrew Macdonald [a], Evan Smal [a], Matt Cross [a], Apeksha Budhkar [a], Kaizad Raimalwala [a], Becca Bonham-Carter [a], Michele Faragalli [a]

[a] *Mission Control Space Services Inc., 162 Elm St. West Ottawa, Ontario, Canada K1R 6N5*
rohaan@missioncontrolspaceservices.com
* Corresponding Author

Abstract

Future space missions will require enhanced perception and autonomy capabilities in the ground and flight segment for complex operations, intelligently process increasingly large volumes of sensor data, and make decisions independent of direct human oversight. Recent advances in deep learning for terrestrial applications have demonstrated that algorithmic approaches can provide a new level of perceptual understanding and data processing capability. The development and deployment of deep learning algorithms for space missions requires a holistic approach to user needs, data curation, model documentation, and production on flight hardware. This paper presents a deep learning framework for space, supported by Mission Control's Spacefarer AI product line, that methodically captures these elements and presents two examples of its use for space flight missions; 1) reprogramming a neural network deep learning model onboard a satellite in low-earth orbit, and 2) the first demonstration of deep learning on the lunar surface.

Keywords: Artificial Intelligence, Deep Learning, EdgeAI, Software Deployment, On-Board Computers

Acronyms/Abbreviations

Spacefarer AI Deployment Toolkit (DT), Central Processing Units (CPUs), Field Programmable Gate Arrays (FPGAS), Graphic Processing Units (GPUS), Visual Processing Units (VPUS), Tensor Processing Units (TPUS), Application-Specific Integrated Circuits (ASICS), Size, Weight, And Power (SWAP), Emirates Lunar Mission (ELM), Neural Network Exchange Format (NNEF), Systems-on-a-Chip (SoC).

1. Introduction

Deep learning techniques are the leading standard in terrestrial applications that require computer vision and natural language processing [1]. In the space sector, Earth observation entities, both government and commercial, rely on deep learning to analyze satellite imagery [2] and in Earth orbit ESA has demonstrated deep learning as part of the flight segment of Φ-sat-1 [3]. Adopting deep learning for space operations is challenging due to the risk tolerance of mission critical applications, the difficulties of risk quantification for statistical techniques, and the extreme environmental requirements imposed on flight hardware; however, owing to increasing commercial and national defence interests, humanity will be conducting more complex operations in space that will require more complex and autonomous operations [4, 5]. Deep learning techniques stand to be a key enabler of both mission autonomy in the flight segment and intelligent decision support in the ground segment as we begin this next phase of planetary exploration.

2. Deep Learning for Autonomy in Space Operations

There are several emerging benefits of using deep learning for autonomous space operations, both in the flight and ground segments. These include:

- Increased Operations Efficiency: Deep learning algorithms can analyze vast amounts of data in real-time, enabling faster decision-making and increased efficiency in space missions.
- Improved Image Processing: Deep learning algorithms are the current State-of-the-Art for image processing tasks such as target recognition, object detection, and image segmentation. These are especially useful in missions involving on-board perception, such as Remote Sensing and Earth Observation satellites, On-Orbit Satellite Servicing missions, and Planetary Robotics.
- Anomaly Detection: Deep learning algorithms can be used to detect anomalies in large datasets, such as sensor readings and telemetry data, which can be used to identify potential issues in real-time, leading to improved mission safety and reliability.

- Autonomous Operations: Deep learning algorithms can be used to enable autonomous operations in space, such as autonomous rendezvous and docking, autonomous landing and take-off, and autonomous spacecraft manoeuvring.

Deep learning techniques use non-linear feature extraction to create an embedded feature space reflecting the properties of the statistical distribution on which they were trained. This means that the underlying robustness and stability of deep learning-based mission software depends on both the data and the code, from the high-level proof of concept model to the executable on the embedded flight hardware. This shift represents a fundamental change in the software design process, from deterministic but limited programming to more universal but data-dependent learning. The need to document both the data and the software, including data provenance and data cleaning, as well as the infrastructure to support continuous deployment and integration, creates new types of technical debt [6]. This requires a change in how software is conceived and tested for spaceflight missions. In this work we introduce an Autonomy Products Pipeline that is used to document and verify the performance of deep learning algorithms for use in spaceflight missions. We describe the key steps involved in the pipeline, explain how it is being used to support two ongoing flight deployments, and conclude with future opportunities for deep learning in space missions.

3. Deep Learning Framework for Space support by SpacefarerAI

To successfully overcome the challenges to deploying deep learning techniques for space operations we must consider the integrity and robustness of the entire data pipeline, from raw data sources to production models. At Mission Control we have developed a deep learning data framework, illustrated in Fig. 1, supported by our AI-specific product-line called SpacefarerAI. The framework provides quality assurance and controls on both data and models inspired by best practices established in terrestrial applications. This includes documentation of the strengths and weaknesses of different training datasets [7] and model reporting [8] that creates a broadly accessible reference of the intended use case for each deployed deep learning model. Human-machine teaming between earth-based users and the autonomous elements of the space segment are core to the framework design. The pipeline starts at the problem definition phase, where we translate user needs into data requirements [9] and create data and model documentation that meets the needs of existing mission program lifecycles, such as critical design reviews and verification and validation activities. We use this pipeline in the development and testing of specific models, such as terrain and novelty classifiers [10], but also in developing software that reduces barriers to entry for development, testing, and integration of algorithms for flight software.

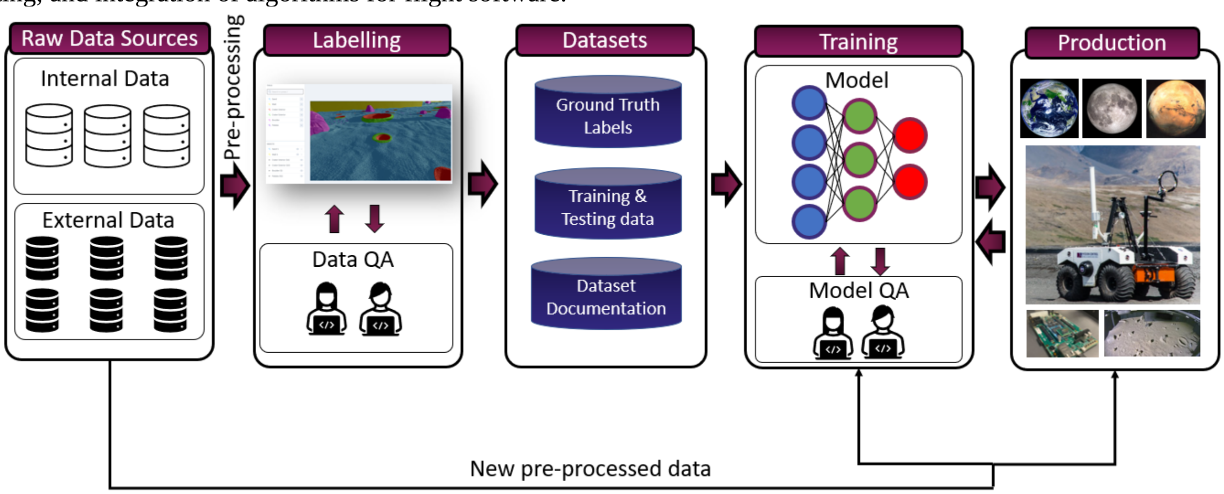

Fig. 1. The deep learning data processing and quality assurance pipeline spans data collection from raw sources to trained production models deployed on spacecraft.

Fig. 1 enforces documentation of the data collection and curation process alongside software engineering effort needed to train and deploy models. It is based around best practices in the field, as defined by the People in AI Guidebook [11] and tools like the AI Reproducibility Checklist [12] and that align the use of deep learning with the intended problem space and outcome in systematic way. The pipeline elements of raw data sources, labelling, datasets, training, and production are general enough to be deployed across a wide array of mission types, from planetary rovers to earth observation satellites, while still allowing customization to specific use cases.

3.1. Raw Data Sources

Mission Control has the ability to generate custom datasets or use established third-party datasets to design and train new deep learning models.

3.1.1. Generating Datasets using Laboratories

Mission Control can generate datasets from sensors deployed in analogue environments [13] for a variety of mission-types using laboratory and field deployments. For lunar robotics, for example, Mission Control uses of our 4000 square foot lunar analogue environment called "the Moonyard" to generate and collect high visual-fidelity datasets. Similarly, our Orbital Autonomy Lab, which consists of two Six Degrees-of-Freedom Robotic Arms, can be used to generate high visual-fidelity and telemetry data for satellite operations, including Rendezvous, Proximity Operations, and Docking/Capture.

3.1.2. External Datasets

Data can also be ingested from external sources, such as large datasets for terrestrial applications like ImageNet [14], Cityscapes [15], or COCO [16] or from planetary science repositories like NASA's Planetary Data System [17] or ESA's Planetary Science Archive [18].

3.2. Labelling and Quality Control

Raw data sources must be pre-processed and down-selected to reflect the needs of the problem space and choice of learning style (e.g., supervised, semi-supervised, unsupervised). The problem space can span multiple sensor types, applications of deep learning from computer vision to optimizing communications bandwidth, and different space applications from satellite remote sensing to space domain awareness to planetary rover navigation. A key element of the data sources and labelling aspects is close consultation between subject matter experts, deep learning practitioners, and end users. This interaction is used to determine the problem ontology, leading to organization of the pre-processed data structure in labelled taxonomies, which are then used teach the compute agent about the problem space. The iterative refinement of the data in this step is a time intensive process but key to reproducing reliable, robust, and trustworthy end results. Automated and manual quality assurance tools and techniques are employed to ensure data and label integrity, screening for alignment between data and higher-level concepts with relevant experts, flagging corrupted labels and mistakes by human operators, and analyzing the properties of the underlying data distribution such as class balance and anomalies.

3.3. Datasets and documentation

Once a dataset has been formalized it may be used to train many different types of models. The core dataset needs to be documented so that users who are many steps removed from the original raw data sources and labelling campaign can understand the structure of the dataset and its intended use cases. We structure dataset documentation using the template provided by Gebru et al. [7], which provides a traceable and accessible point of reference for all future uses of that data in model training and production deployment. A dataset datasheet can also be versioned in releases, as future missions may incorporate new data or dictate the need to combine previous datasets into one.

3.4. Model Training and Quality Assurance

Model training occurs using open-source high level languages designed for deep learning on cloud or local compute infrastructure. The process of training a model and tuning hyperparameters is documented using automated tools to track performance. High-level models span a wide number of domains and sensor types: feature detection, anomaly detection, image classification, depth estimation, visual odometry, terrain classification, and more. Additionally, one model may have several different use cases, especially if it is released externally. Once a tuned model has been approved and verified for release, we create a Model Card [8], as illustrated in Fig. 2, to capture a broadly accessible description, intended use, performance metrics, source training data, recommendations for use, and any ethical considerations in using the model.

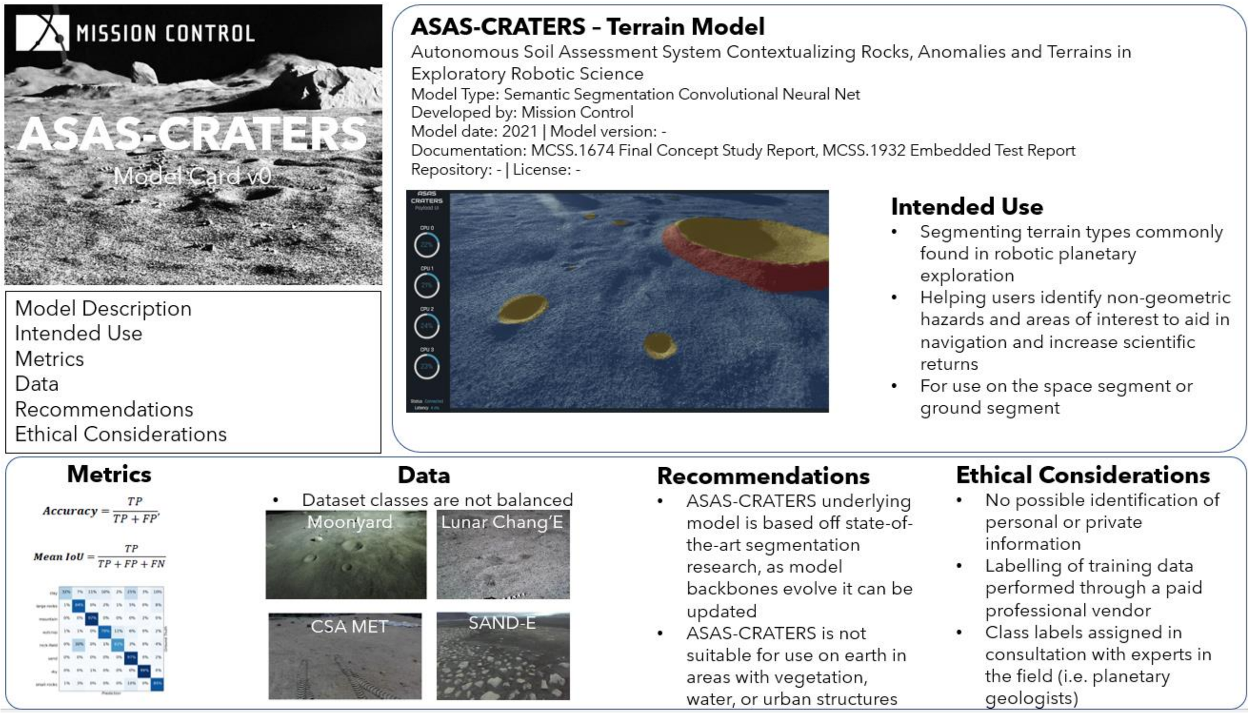

Fig. 2. Model cards, such as the one for this semantic segmentation terrain classifier, are used to document the model performance and intended use cases for experts and non-experts in deep learning.

3.5. Production Deployment

Production deployment of trained models on space flight hardware is a bottleneck to wider adoption of deep learning within space missions. We have developed a multi-stage, deep learning compiler and corresponding run-time for accelerating deep learning inference, called the Deep Space Compiler (DSC) and Deep Space Runtime (DSR), respectively. This Spacefarer AI Deployment Toolkit (DT) is a software product that takes pre-trained machine learning models built in a high-level format and implements them on low power systems-on-a-chip (SoC) suitable for edge compute applications in space, through the steps illustrated in Fig. 3. The DT addresses the difficulty of implementing deep learning models in a landscape of increasingly heterogeneous compute architectures that includes central processing units (CPUs), field programmable gate arrays (FPGAs), graphic processing units (GPUs), visual processing units (VPUs), tensor processing units (TPUs), and application-specific integrated circuits (ASICs) [19]. It is tailored to the specific needs of edge computing devices in space that perform deep learning [20], in which low size, weight, and power (SWaP) requirements are combined with radiation hardening and robustness to thermal cycling and vacuum conditions.

As an intermediate format we rely on the Neural Network Exchange Format (NNEF) [21]. The DSC consumes the model in NNEF and lowers it to an abstract representation where semantic checking and optimizations can be performed on deep learning architecture. Fig. 3 illustrates this process for a model trained in a high-level framework with outputs from packages like Tensorflow or PyTorch, which may be in their native format or converted to an intermediate standard like ONNX [22]. The NNEF tools ingest and convert these formats and sends input the resulting .nnef file to the compiler. The DSC outputs a bytecode executable that can then run the model on the edge device with new inputs in flight.

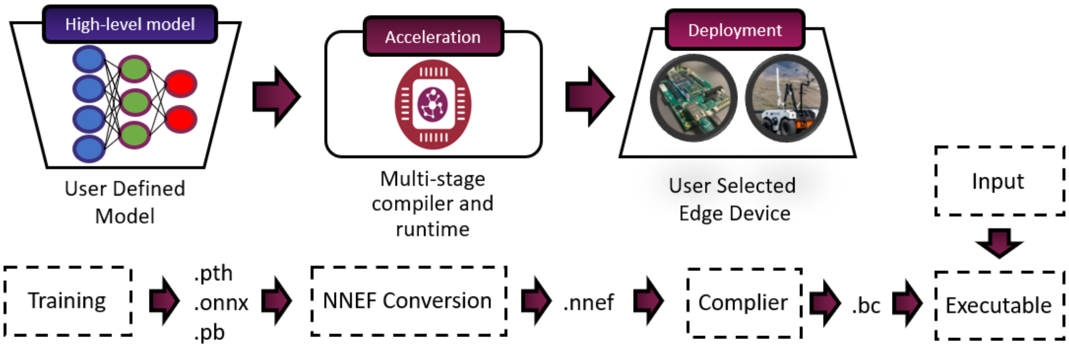

Fig. 3. The Spacefarer AI Deployment Toolkit accelerates high-level trained models into production of edge devices for space flight applications, moving from training to compiling to running the model.

In the following section we will explore two examples of the Autonomy Products Pipeline and DT in the deployment of deep learning algorithms onboard flight missions in 2023. This includes the first use of deep learning on the lunar surface and gaining experience in optimizing deep learning run-times for use in low earth orbit. These missions will inform better data practices and provide insights into new validation techniques in an iterative cycle to continually improve mission performance.

4. Autonomy Deployment in Flight Missions

Modern deep learning has the potential to help space operations to become more autonomous, adapt to unexpected changes, synthesize data for human analysis, and maximize productivity during idle time. Mission Control has already shown the value of deep learning in an analogue planetary exploration mission [13] and in 2023 will demonstrate this potential in two space missions. Our deep learning terrain classifier will be deployed onboard the ispace lunar lander mission M1 and will directly support the Emirates Lunar Mission (ELM) micro-rover with the first capability demonstration of artificial intelligence on the lunar surface. The DT will also be used to benchmark the performance of neural network models running on an FPGA on ESA's OPS-SAT for an Earth Observation application.

4.1. Lunar Artificial Intelligence Capability Demonstration

Rapid characterization of visual imagery is a necessary step to enable the planning and execution of navigation and science objectives for a planetary rover mission [10]. The ability to perform characterization of a planetary surface autonomously and automatically is useful to accelerate decision making and circumvent communication or compute constraints imposed by operating on another planetary body. Mission Control has used its deep learning framework to design and test terrain classification models for this purpose using convolutional neural networks trained to perform semantic segmentation. The raw data source has been curated from analogue environments in the field [13] and Mission Control's Moonyard, with training imagery shown in Fig. 4.

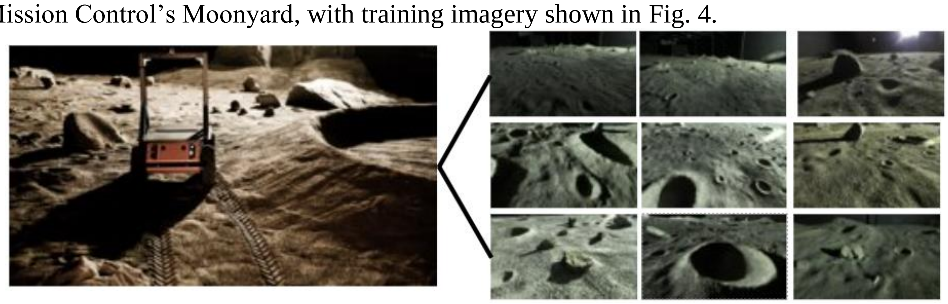

Fig. 4. Training data gathered using Mission Control's Moonyard, demonstrating variation in geological features and lighting conditions reflective of the lunar surface.

Labelling is performed in conjunction with experts in planetary geology. This leads to a class taxonomy that captures science targets and visually distinctive features of the terrain. Labels are generated at the pixel level and checked for integrity and consistency by humans-in-the-loop. The resulting dataset is documented with a datasheet and consists of Moonyard imagery and ground-truth masks. Fig. 5 illustrates the steps once the curated dataset and high-level model training are underway, with input imagery and ground-truth masks used to verify model performance and generate prediction overlays on the original image. Iterative training using a separate validation and test set allows optimization of the model hyperparameters.

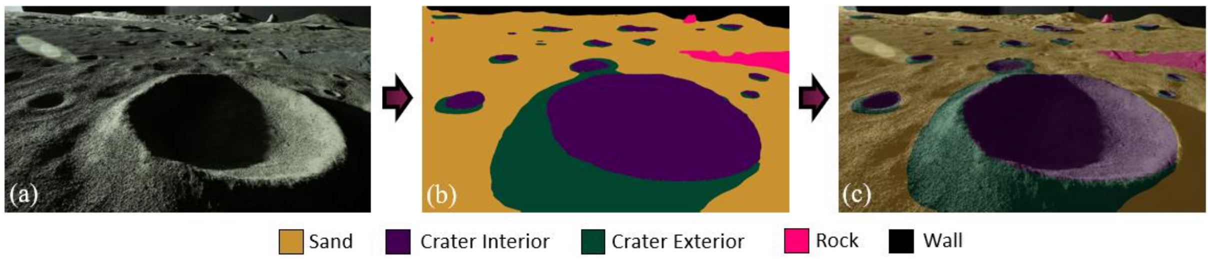

Fig. 5. Steps of the terrain classifier model training and validation (a) input image taken in the Moonyard, (b) ground-truth mask and (c) model prediction overlaid on input image providing rapid segmentation of sand, craters, rocks, and the background wall.

Of the set of trained models, the model that achieves the highest performance both in terms of accuracy and speed is selected for implementation on the flight hardware. The DT is used to migrate the high-level model to a specific SoC, in the case of the lunar AI capability demonstration this is a Xiphos Q7S with an ARM-based CPU and Xilinx FPGA. The model must pass integration tests with the mission software and hardware to ensure the power and memory requirements and compatible with the limitations of the hardware. Unit tests ensure that the model output and performance are not degraded by running pixelwise comparisons of the output between the high-level model and the embedded model. The implemented model, MoonNet, is described in a model card as per the framework's documentation requirements. MoonNet will be the first deep learning algorithm to be used on the lunar surface. The mission launched on 11 December 2022 and is staged for lunar operations in the Second Quarter of 2023.

A ground-based version of MoonNet is used to derive value-added products from the semantic maps of the scene. The ability to perceive features of interest without *a priori* knowledge of the landing site on the moon or the placement of specific features allows for the creation of data products that highlight and augment information for the mission science team. Fig. 6 shows an example of one such product derived from the MoonNet outputs. Fig. 6 (a) shows an input image from the Moonyard containing sand, craters, rocks, and the background wall. Science users may have mission objectives tied specifically to craters and rocks on the lunar surface and so the segmentation map prediction is used to generate a smaller image product that keeps the high-resolution of these features while down sampling the background wall and foreground sand. In future missions these kinds of derived data products can be constructed onboard and used to maximize the science return in limited communications bandwidth scenarios.

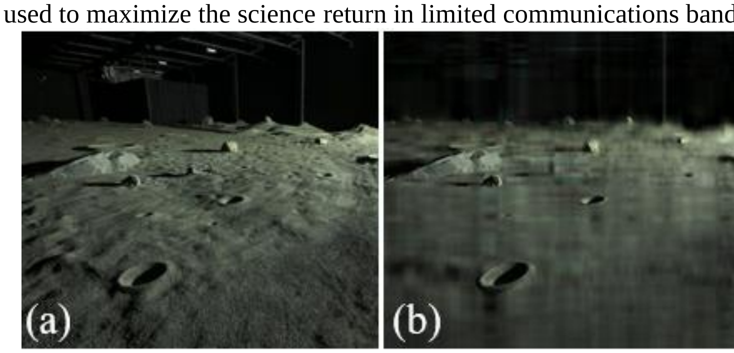

Fig. 6. Example of a derived data product from the MoonNet terrain classifier predictions, the output prediction is used to selectively down sample certain image features to create a smaller image while maintaining high fidelity for craters and rocks.

4.2. AI Demonstration in Low Earth Orbit on ESA OPS-SAT

The deep learning framework applies across a variety of mission classes and edge compute types. To demonstrate its effectiveness, Mission Control is executing a comparative study of high-level and low-level implementations of deep learning models for remote sensing satellites using ESA's OPS-SAT experimental platform. The SmartCam model [24] onboard runs on an Intel Cyclone V SoC. The DT is used to re-deploy the model on the SoC FPGA, and then compare performance between the CPU implementation currently in use and the FPGA version, as illustrated in Fig. 7.

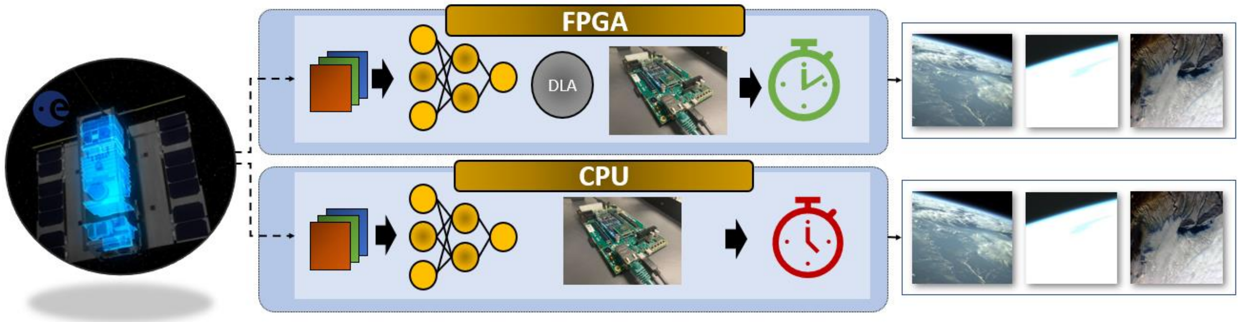

Fig. 7. OPS-SAT imagery is processed onboard using an FPGA and CPU based implementation of a neural network image classification model (SmartCam), Mission Control's Spacefaer AI Deployment Toolkit is used to deploy the model on the FPGA.

FPGAs are used in spacecraft from orbital satellites to rovers to space stations. The ability to reprogram an on-orbit FPGA to deploy deep learning models using the DT will demonstrate a new capability, expanding the potential for the deployment of deep learning from spacecraft currently in the design phase to existing spacecraft with hardware that may not have been intended to include deep learning applications in their initial design.

5. Conclusion

The ability to deploy validated and trustworthy deep learning software within embedded space system hardware will have significant impacts on space mission capabilities going forward. Increased autonomy allows for new types of missions and greater ability to achieve commercial, defence, and scientific objectives. This is only possible if the development approach is made transparent, accessible, and reproducible to mission stakeholders. In this work we have presented a deep learning framework that captures and documents the steps from raw data selection to onboard software testing and deployment. In addition to software engineering code documentation, we advance the need for documentation at the dataset and model level, to ensure that trained models meet the intended use case, and it is understood by end users their domain of applicability and where they may encounter out-of-distribution examples during flight. We demonstrate our use of this framework for two flight missions, in the development of the training dataset and MoonNet lunar terrain classifier for the first demonstration of artificial intelligence on the lunar surface and for deployment of the SmartCam model onboard the FPGA of a satellite SoC payload.

With two flight missions gaining space flight heritage in the next year, deep learning on edge devices in space will continue to advance. Future developments will focus on developing deep learning tools that can interface with mission critical systems, sustained operations that update model weights online, and new approaches inspired by fundamental research in AI. As edge devices proliferate, new approaches such as Federated Learning can update deep learning models on aggregate without the need to send large amounts of raw data back to ground stations on Earth. At the same time, derived data products that use deep learning as a processing step can help end users and operators to unlock new insights, setting up the possibility of Human-AI teams working at the intersection of flight and ground segment.

Acknowledgments

Mission Control acknowledges the sponsorship of the Canadian Space Agency in developing the ASAS-CRATERS terrain classification work and Lunar AI Capability Demo through its STDP and LEAP programs. Mission Control would also like to acknowledge the European Space Agency's support on the OPS-SAT project.

References

[1] Zhang, D, Maslej N., Brynjolfsson, J, Etchemendy J., Lyons, T., Manyika J., Ngo H., Niebles, J.-C., Sellitto, M, Sakhaee E, Shoham, Y., Clark J., and Perrault, R., "The AI Index 2022 Annual Report," *AI Index Steering Committee*, Stanford Institute for Human-Centered AI, Stanford University, March 2022.

[2] Ball J. E., Anderson D. T., and Chan Sr. C. S., "Comprehensive survey of deep learning in remote sensing: theories, tools, and challenges for the community", *Journal of Applied Remote Sensing* 11(4) 042609 (2017).

doi: 10.1117/1.JRS.11.042609

[3] Giuffrida G., Fanucci L., Meoni, G., Batič, M., Buckley, L, Dunne, A., van Dijk, C., Esposito, M., Hefele, J., Vercruyssen, N., Furano, F., Pastena, M., and Aschbacher J. (2021) "The Φ-Sat-1 Mission: The First On-Board Deep Neural Network Demonstrator for Satellite Earth Observation," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-14, Art no. 5517414
doi: 10.1109/TGRS.2021.3125567

[4] International Space Exploration Coordination Group (ISECG), "The Global Exploration Roadmap 3rd edition," NP-2018-01-2502-HQ (2018)

[5] National Aeronautics and Space Administration (NASA) (2020), "Artemis Plan NASA's Lunar Exploration Program Overview," NP-2020-05-2853-HQ

[6] D. Scully, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhaury, M. Young, J.-F. Crespo, and D. Dennison. "Hidden Technical Debt in Machine Learning Systems", *2015 Proceedings of the Neural Information Processing Systems Conference.* Vol. 2 Pages 2503-2511, December 2015

[7] Gebru T., Morgenstern, J., Vecchione, B., Vaughan, J., Wallach, H., Daumé III, H., and Crawford K., "Datasets for Datasheets" *Communications of the ACM, December 2021, Vol. 64 No. 12, Pages 86-92* (2018)
doi: 10.1145/3458723

[8] Mitchell M., Wu S., Zaldivar A., Wu, S., Zaldivar A., Barnes P., Vasserman L., Hutchinson B., Spitzer E., Raji I., and Gebru, T. "Model Cards for Model Reporting*," FAT*'19 Conference on Fairness, Accountability, and Transparency,* pg. 220-229 (2019)
doi: 10.1145/3287560.3287596

[9] Google PAIR, "People in AI Guidebook," *User Needs + Defining Success* pair.withgoogle.com/guidebook Published May 8, 2019, Updated May 18, 2021 (2019)

[10] Raimalwala K. V., et al., "Automated Characterization of Planetary Surface Geological Features from Surface Spacecraft Cameras to Support Mission Operations", 5th Planetary Data and PSIDA LPI Contribution No. 2549 2021

[11] Google PAIR. "People in AI Guidebook." pair.withgoogle.com/guidebook Published May 8, 2019, Updated May 18, 2021

[12] Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., Fox E., and Larcholle, H., "Improving Reproducibility in Machine Learning Research," *2019 Proceedings of the Neural Information Processing Systems Conference (NeurIPS)*

[13] Sinha, P., Horgan B. H. N., Ewing R. C., Rampe E. A., Lapotre M. G. A., Nachon M., Thorpe M., Cedford C., Mason K., Champion E., Gray P. C., Reid, E., and Faragalli M., "SAND-E: A Rover-Based Mars Analog Study of a Mafic Sedimentary Environment," *Seventh Mars Polar Science Conference* LPI Contribution 2099 2020

[14] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). "Imagenet: A large-scale hierarchical image database," *In 2009 IEEE conference on computer vision and pattern recognition* pp. 248–255 (2009).
doi: 10.1109/CVPR.2009.5206848

[15] Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Beneson, R., Franke, U., Roth, S., and Shiele, B.. "The Cityscapes Dataset for Semantic Urban Scene Understanding," *2016 Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR).*
doi: 10.1109/CVPR.2016.350

[16] Lin, T.-Y., Maire M., Belongie S., Bourdev, L., Girshick R., Hayes J., Perona, P., Romanan, D., Zitnick, C. L., and Dollár, P. D. "Microsoft COCO: Common Objects in Context," Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. (2015)
doi: 0.1007/978-3-319-10602-1_48

[17] NASA Planetary Data System, last updated March 2022 https://pds.nasa.gov/

[18] ESA Planetary Data Archive, 2022 European Space Agency https://open.esa.int/esa-planetary-science-archive/

[19] Reuther, A., Michaleas P., Jones, M., Gadepally V., Samsi, S., and Kepner J. "Survey and Benchmarking of Machine Learning Accelerators," 2019 IEEE High Performance Extreme Computing Conference (HPEC), 2019, pp. 1-9,
doi: 10.1109/HPEC.2019.8916327.

[20] Furano G., Meoni G., Dunne A, Moloney D., Ferlet-Cavrois, Tavoulais, A., Byrne, J., Buckley L., Psarakis M., Voss K.-O., and Fanucci L. "Towards the Use of Artificial Intelligence on the Edge in Space Systems: Challenges and Opportunities," in IEEE Aerospace and Electronic Systems Magazine, vol. 35, no. 12, pp. 44-56, 1 Dec. 2020,
doi: 10.1109/MAES.2020.3008468

[21] Khronos Group Neural Network Exchange Format (NNEF) NNEF Overview - The Khronos Group Inc, 2022

[22] Labrèche G., Evans, D., Marszk D., Mladenov T., Shiradhonkar V., Soto, T., and Zelenevskiy, V., "OPS-SAT Spacecraft Autonomy with TensorFlow Lite, Unsupervised Learning, and Online Machine Learning," *in 2022 IEEE Aerospace Conference* 2022