SSC25-IX-06

Development and Test of Deep Learning Techniques for Stereo-based Pose Estimation of Small Satellites

Tim Heydrich, Becca Bonham-Carter, Apeksha Budhkar[1], Hugo Burd[1], Galen O'Shea[1], Luis Chavier, Adam Deslauriers, Alan Higginson, Sharanjit Virdi[1], Andrew Macdonald, Kaizad Raimalwala, Michele Faragalli
Mission Control
162 Elm St. W., Ottawa, ON;
tim@missioncontrolspace.com

ABSTRACT

For in-orbit servicing, spacecraft must conduct a variety of tasks known as Rendezvous & Proximity Operations, Docking (RPOD). The ability to accurately and reliably estimate a target object's pose is required to perform these tasks. In computer vision, deep learning techniques been shown to significantly outperform classical techniques. However, there are several barriers to adoption of such technologies in spaceflight, related to development and verification/validation processes, and constraints of the target flight hardware.

In this paper, we outline our approach and limited results from a recently completed project on the development and test of a neural network designed to allow a servicing spacecraft to estimate the pose of a target object in proximity, in real-time, using stereo vision, so as to enable planning for subsequent maneuvers.

Developing and testing pose estimation approaches requires a reliable test bed capable of providing representative simulation of relative movements as well as hardware analogous to that used for space applications. To address this, we first set up a facility we named the Orbital Autonomy Lab (OAL), which consists of two Universal Robots arms and an OptiTrack motion capture system. One arm was configured as a servicing spacecraft arm with a stereocamera attached; the other was used to place a mock satellite in several pose configurations, in view of the stereocamera. Using this setup, we collected 11,539 stereo pairs of a mock satellite in different positions and lighting configurations, emulating realistic approach and inspection paths.

A core facet of this study was to gauge performance of neural networks on a flight-representative board. Given our previous experience in deploying CNNs for actual spaceflight missions, the Kria KV260 was selected as it is very similar to the popular Xiphos Q8 that was used in previous projects.

For model development, we considered key constraints. First, the model must be lightweight to allow for quick inference on a flight-rated board. Second, the model needs to learn the target's pose without prior knowledge of its shape. Third, the model needs to be generalizable across target poses and lighting conditions and approach scenarios. Following our previous experience in developing neural networks for spaceflight applications, we investigated many design approaches. During testing, we tracked L1 loss, orientation error, translation error, and inference speed, to gauge performance. Subsequently, we developed and tested three models, and selected one, dubbed PosenatorV2 with a dual MobileNetV3 backbone encoder as a final candidate for deployment on the KV260 for testing.

Our targets for Mean Translation Error and Mean Angular Error were 0.5cm and 5° respectively. Through our final test results for the dual-arm sequences, we were able to achieve 0.61cm and 2.34° respectively. For assessing performance on the Kria KV260, we were able to achieve a time from image capture to inference at just under 1.5 seconds. In comparison, running this experiment with a model on a Heroku server, we were able to achieve a time of ~0.5 seconds.

INTRODUCTION

For in-orbit servicing, spacecraft must conduct a variety of operations within the umbrella of Rendezvous & Proximity Operations, Docking (RPOD). The ability to accurately and reliably estimate a target object's pose is

[1] No longer employed at Mission Control.

required for the servicing spacecraft to perform subsequent motion planning and execution.

In computer vision, deep learning techniques been shown to significantly outperform classical techniques. However, there are several barriers to adoption of such technologies in spaceflight, related to development and verification/validation processes, as well as technological constraints of the target flight hardware. Mission Control's AI & Autonomy team focuses on addressing these barriers and making it feasible to leverage deep learning technologies for in-flight applications ranging from robotics to Earth Observation.

In previous work, Mission Control has had experience in two spaceflight missions. In 2023, Mission Control's MoonNet CNN model deployed on a Xiphos Q7 flew to the Moon as a technology demonstration payload onboard the first mission by Japanese company ispace inc [1]. Also in 2023, Mission Control performed an experiment to re-deploy the publicly available SmartCam model on ESA's OPS-SAT cubesat platform, using our proprietary deployment tools [2].

In this paper, we outline our approach and limited results from a recently completed project (2023-2024) on the development and test of a neural network designed to allow a servicing spacecraft to estimate the pose of a target object in proximity, in real-time, using stereo vision, so as to enable planning for subsequent maneuvers.

PROBLEM STATEMENT

Estimation of a target object's pose is a fundamental component in a holistic solution for autonomous RPOD tasks. With a validated pose estimate, a servicing spacecraft can perform subsequent trajectory planning and execution, inspection and capture tasks, among others. This is a broadly applicable capability required for scenarios ranging from spacecraft assembly to spacecraft servicing, debris rendezvous and capture, asteroid rendezvous and landing, among others. Thus, a high-performance pose estimation solution for spacecraft operating in complex and unstructured environments would be a transformative solution that servicing companies can incorporate in their system, which in turn enables them to further produce more innovative solutions built on this technology.

The key problem statement and investigation that was considered in this study was to identify whether model-free pose estimation techniques based on stereoscopic optical vision be trusted to perform at a level of reliability appropriate for use in autonomous/semi-autonomous operational environments.

Secondary problem statements were also noted for consideration:

- Are AI-based computer vision techniques for RPOD applications suitable for low-powered space computers?
- Can laboratory facilities generate high fidelity visual data sets for training & testing RPOD systems?

SOLUTION APPROACH

Deep learning-based techniques have dominated the field of computer vision in recent times since the appearance of large-scale deep neural networks. The winners of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) transitioned from classical to deep learning based methods since landmark work by Krizhevsky, Hinton et al. in 2012 [3], [4].

AI-based pose estimation techniques that use deep learning achieve a higher level of performance than traditional methodologies, given their superior abilities in identifying features and patterns that regular, hand-engineered pose-estimation algorithms may not detect. This has been demonstrated recently by researchers exploring the use of CNNs for pose-estimation in various rendezvous and proximity operations [5], [6]. Our hypothesis around suitability of these algorithms for on-board processing on space-rated hardware is supported by our own extensive experience in developing AI computer vision models for use in resource constrained spaceflight applications [7].

AI-based pose estimation techniques for RPOD operations can achieve a higher level of performance in pattern detection compared to classical techniques. As such, using AI-based algorithms accommodate a broader set of use cases and ConOps, which enables a diversity of mission types to be accomplished in challenging and dynamic environments, ranging from the orbital assembly, debris mitigation, and satellite refueling.

One key aspect of our approach was training a model to be capable of pose estimation of a fully unknown target. Previous approaches have relied on some amount of existing knowledge, for example key points, specific features and their position with respect to each other on the target, or a complete 3D model. We wanted our model to be versatile and not be constrained to only known targets.

To address the problem statements highlighted earlier, our goal was to develop a proof-of-concept solution that is feasible to run on a flight-representative embedded system for real-time applications in HITL (Hardware In The Loop) fashion. To develop a robust and well-performing solution that can reliably predict pose for different targets in different scenarios was not in scope for this project. For example, while the true test of a generalizable pose estimation solution for multiple targets would also include a range of deformed debris-like objects, for simplicity purposes we only devised two types of satellite-shaped objects for this project.

TEST SETUP: ORBITAL AUTONOMY LAB

Orbital Autonomy Lab

Developing and testing pose estimation approaches requires a reliable test bed capable of providing representative simulation of relative movements as well as hardware analogous to that used for space applications. To address this, we first set up a facility we named the Orbital Autonomy Lab (OAL), which consists of two Universal Robots arms, an OptiTrack motion capture system, and two mock satellites. A system that partly represented the Lunar Gateway was also set up for the purposes of an educational experience delivery program. One arm, dubbed *Atlas*, was configured as a servicing spacecraft arm with a stereocamera attached; the other, dubbed *Ada* was used to place a mock satellite in several pose configurations, in view of the servicing arm's stereocamera. Using this setup, we collected 11,539 stereo pairs of different mock satellites in different positions and lighting configurations, emulating realistic approach and inspection paths. Five types of sequences were used for data collection.

Mission Control's Spacefarer^{TM}, a commercially available product for space robotics operations, was used to develop a laboratory operations system with user interfaces for arm control, data visualization, and other functions.

Hand Eye Calibration

In order to determine ground-truth pose when collecting data the servicer arm needs to know the relative position to the gripper on the target arm. Two different approaches were used, the initial, utilized Brown-Conrady distortion model [8] with a checkerboard mounted to the target arm. The series of calibrations were stereocamera to servicer gripper, servicer arm to target gripper and servicer arm base to target arm base.

This basic approach was then refined by the integration of the OptiTrack motion capture system. We still had to determine the transforms for the sterocamera to its enclosure which was tracked by OptiTrack. The other calibrations were no longer required as the objects were being tracked by the motion capture system which allowed us to extract the transforms directly. The system utilizes infrared markers on the camera enclosure and the target to track them in 3D space.

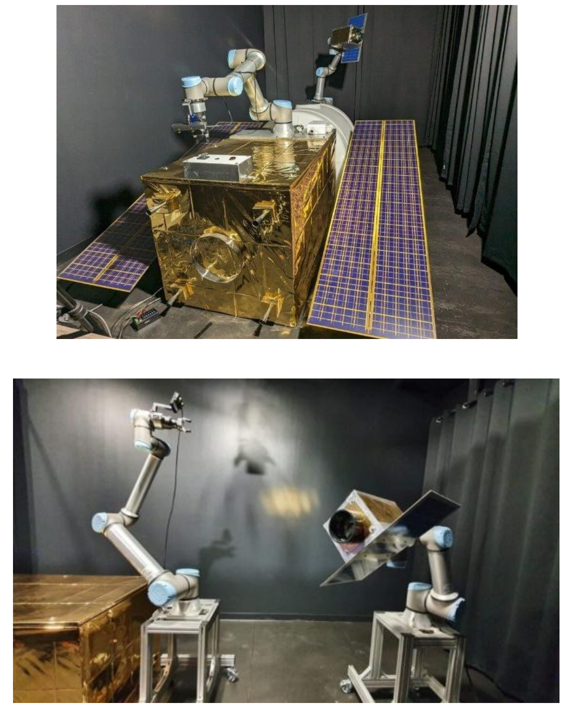

Figure 1: Front-facing and side-facing views of the Orbital Autonomy Lab at Mission Control's HQ.

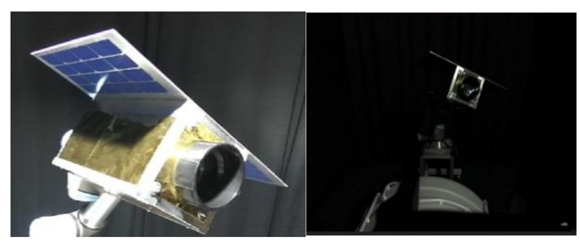

Figure 2: A view of PlywoodSat from two different views in different lighting conditions.

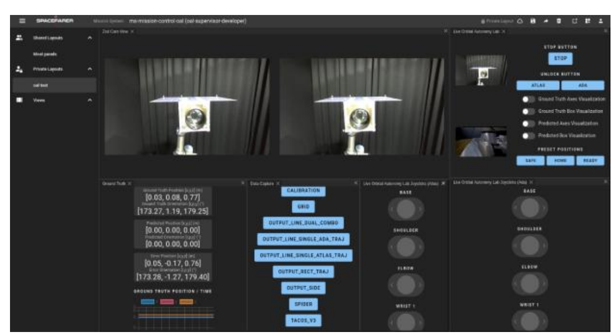

Figure 3: A screenshot of the lab control interface developed with Spacefarer^{TM}.

Dataset Planning and Collection

A critical factor for a high-performance Deep Learning algorithm is a high-quality dataset available for training and validation. There is a need for publicly available datasets that AI researchers can benefit from to advance their work in orbital autonomy scenarios. However, given the nature of autonomous in-orbit servicing applications, generating high-quality datasets in space is a non-trivial task. Today, a majority of the algorithms in the in-orbit servicing field rely on synthetic data generation to generate imagery for in-orbit servicing computer vision applications. Current state-of-the-art synthetic datasets available for public use include ESA's SPEED Pose Estimation Dataset for monocular imagery, Unreal Rendered Spacecraft On-Orbit (URSO), and SPARK (SPAcecraft Recognition leveraging Knowledge of space environment), which is a multimodal dataset with synthetic labeled imagery of spacecraft. One of the drawbacks of using synthetic imagery is that neural networks are not invariant to environmental changes, thus a network's accuracy on a synthetic dataset is not guaranteed to be maintained when tested in real-world conditions [5]. This is shown by Park et. al. when testing their state-of-the-art SPN Pose Estimation Network trained on synthetic datasets. While Park uses a Neural Style Transfer (NST) technique to mitigate network performance loss as a result of training on synthetic data, there are technological unknowns surrounding the robustness of those techniques.

To accomplish the objectives of training and testing a neural network model for pose estimation, we developed a plan to generate a novel dataset of stereoscopic satellite imagery captured in the OAL, geared towards applications critical to RPOD, such as segmentation and pose-estimation. The dataset was documented using Datasheets for Datasets [9] and verified for data quality in accordance with our best practices around machine learning data.

Before collecting any data, we planned different likely maneuvers that a servicer craft would take when approaching or inspecting a target. We devised two main paths that we wanted to collect data for: 1) "Approach" where the servicer would slowly move towards the target either in a straight line or an arc, and 2) "Inspection" where the servicer would circle around the target to visualize all sides. Initially, we broke these two into smaller sub-collections, characterized by different approach angles, approach distances, having the target remain stationary or rotate. Ultimately, we performed one holistic dataset collection which we called "dual-arm" sequence, which combined the Approach and Inspection as well as performed a rotation match with the target followed by a final approach to dock.

In addition to the different satellite positions and collection sequences, we devised a series of repeatable lighting setups that emulate the harsh lighting conditions in orbit. We selected four different lighting setups which all had different incident angles, intensity and emulated albedo from the Earth.

Embedded Target

A core facet of this study was to gauge performance of different neural networks on a board that could be considered representative of flight-rated boards.

Given previous experience in deploying CNNs for actual spaceflight missions, we selected the Kria KV260 as it is very similar to the popular Xiphos Q8 that we have previously worked with. Similar to the Q8, the KV260 uses a Xilinx Ultrascale FPGA architecture, and uses an ARM Cortex-A53 and Application and Arm Cortex-R5F Real Time processor. To facilitate experiments, we set up the appropriate software and toolchains to interact with the neural network models and the OAL.

Figure 4: The Kria KV260 board

NEURAL NETWORK DEVELOPMENT

To train the neural network to perform relative pose estimation from stereo images, we had to consider the following constraints:

1) The model has to be lightweight enough to allow for quick inference on a flight-rated computer.
2) The model needs to learn the target's pose without prior knowledge of its shape
3) The model needs to be generalizable across target poses and lighting conditions, trained on data representative of real-world scenarios.

Following our previous experience in developing neural networks for spaceflight applications, we investigated several design approaches. As we developed and tested different approaches, we tracked L1 loss, orientation error, translation error, and speed score, to gauge performance.

We developed these early machine learning models alongside advancing our OAL capabilities. Advances made with integrating Spacefarer with the OAL and adding enhanced capabilities like motion capture allowed us to refine our machine learning models and collect significant amounts of data.

Model testing involved testing on collections made with different lighting and satellite positions. We tested model performance and inference speed on both our embedded target and a cloud target, both with different model runtimes to compare their results.

Early approaches included testing convolutional residual networks (CRNNs) and temporal convolutional networks (TCNs) but did not yield good results. Subsequently, we developed and tested models we dubbed:

1) StereoNet: a ResNet-18 backbone in a twinned architecture inspired by Siamese networks and Stanford's Speed+ [10] pretraining for network weights. This showed good results in training but was too large for real-time applications.
2) Posenator: a custom dual backbone encoder. This was much smaller in size and complexity but showed worse results than StereoNet.
3) PosenatorV2: a dual MobileNetV3 backbone encoder. This was also much smaller than StereoNet but had better performance than Posenator, and was selected as the final candidate for deployment on the Kria KV260.

Key characteristics of these models are summarized in the table below, the model parameters as well as the number of Multiply Accumulate (MAC) operations.

Table 1. Characteristics of the models

Model Architecture	Parameters	GigaMACs
StereoNet	22,360,299	3.64
Posenator	2,983,485	0.72
PosenatorV2	3,041,863	0.12

RESULTS AND DISCUSSION

Initial training utilized only one type of sequence at a time. Initial results on single sequences were okay but very poor across all images at the time, as models were unable to generalize properly, largely due to insufficient data and the size of the problem space. At this stage we decided to further augment the dataset, notably by also adding a 'grid' sequence for the stereocamera's motion, which expanded the distribution of poses for training. This improved the model's ability to generalize and increase performance.

As PosenatorV2 was trained on sequences where only the arm representing the servicing spacecraft was moving with respect to the target object, we then trained it on more complex sequences where the target object was also made to move using the robotic arm hosting it. The model was able to learn this 'dual-arm' sequence and produce good results in varying lighting conditions.

Our targets for Mean Translation Error and Mean Angular Error were 0.5cm and 5^{o} respectively. Through our final test results for the dual-arm sequences, we were able to achieve 0.61cm and 2.34^{o} respectively.

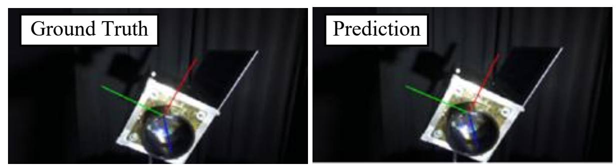

Figure 5: Comparison of a single image and pose: ground-truth vs. pre-diction from PosenatorV2.

FUTURE WORK AND CONCLUSION

In this paper, we presented our approach and limited results from initial development of neural network models designed to allow a servicing spacecraft to estimate the pose of a target object for RPOD tasks.

In future work, we aim to work with government and industry partners to build on this technology and other capabilities to advance the state of the art of RPOD operations for future flight opportunities.

The current model ingests only stereo camera images without any prior knowledge of the target or additional sensor input. Future iterations of the model could either be fine-tuned with additional knowledge of the target, for example keypoints or add sensor fusion to support input from additional onboard sensors like RADAR or LIDAR.

Our current target, *PlywoodSat*, resembles a small CubeSat as we developed a model that works without prior knowledge of the target. To test our model-free approach for more real-world scenarios consisting of RPO with ill-defined or unstructured targets, a future extension of this work would be to add objects that are representative of space debris to our training and test dataset.

Furthermore, additional work needs to be done in ConOps development. Our current proof-of-concept approach did not consider how real-world servicing missions would be performed and what level of input from operators would be required. This work would focus on determining if the satellite independently picks which part of the target would be considered up or which part of the target would be best for docking or grappling. With other approaches that have existing knowledge of the target, this is better defined as more pre-mission planning can be done. However, with our approach this needs to be done after the inspection phase either autonomously or with input from the ground.

Advancing MLOps

This project provided the first opportunity for our team to conduct MLOps on flight-representative hardware, by running and updating models on the Kria board. While PosenatorV2 has only been tested inside our OAL, since the completion of this project, we have made significant advances in model deployment and continuous operations by developing a complete MLOps pipeline that can be utilized for a wide range of mission scenarios. This development has allowed Mission Control's AI & Autonomy team to be able to serve models to fly on multiple missions, including the Mission Persistence satellite platform that is scheduled to launch on SpaceX' Transporter-14, as well as a future lunar rover mission [7]. This pipeline allows us to deploy and test different models and update them in flight as we collect more data during the mission. In these future projects, our team will gain more experience in the design and practice of MLOps that will then benefit our further advancement of ML models for RPO pose estimation.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Canadian Space Agency (CSA) [22STDPS10].

REFERENCES

[1] M. Cross *et al.*, "Enabling Autonomy and Operations for Lunar Surface Missions: An Overview of Demonstrated Capabilities," in *17th Symposium on Advanced Space Technologies in Robotics and Automation*, Leiden, Netherlands, Oct. 2023. Accessed: Apr. 22, 2024. [Online]. Available: https://robotics.estec.esa.int/ASTRA/Astra2023/2023%20PAPERS/70_Cross_Enabling%20Autonomy%20and%20Operations%20for%20Lunar%20Surface%20Missions%20An%20Overview%20of%20Demonstrated%20Capabilities.pdf

[2] L. Chavier *et al.*, "Deploying Artificial Intelligence Capabilities by Hybridizing a Neural Network on a Satellite," presented at the ASCEND, AIAA, 2023. doi: https://doi.org/10.2514/6.2023-4718.

[3] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015, doi: 10.1007/s11263-015-0816-y.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[5] T. H. Park, S. Sharma, and S. D'Amico, "Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft," Sep. 01, 2019, *arXiv*: arXiv:1909.00392. doi: 10.48550/arXiv.1909.00392.

[6] L. P. Cassinis, A. Menicucci, E. Gill, I. Ahrns, and M. Sanchez-Gestido, "On-ground validation of a CNN-based monocular pose estimation system for uncooperative spacecraft: Bridging domain shift in rendezvous scenarios," *Acta Astronaut.*, vol. 196, pp. 123–138, 2022.

[7] B. Bonham-Carter *et al.*, "Advancing AI Reliability for Space Missions through Machine Learning Operations," in *18th International Conference on Space Operations*, Montreal, QC, Canada, May 2025.

[8] D. C. Brown, "Decentering distortion of lenses, Photogrammetric Eng," 1966.

[9] T. Gebru *et al.*, "Datasheets for datasets," *Commun. ACM*, vol. 64, no. 12, pp. 86–92, Dec. 2021, doi: 10.1145/3458723.

[10] T. H. Park, M. Märtens, G. Lecuyer, D. Izzo, and S. D'Amico, "SPEED+: Next-generation dataset for spacecraft pose estimation across domain gap," in *2022 IEEE aerospace conference (AERO)*, IEEE, 2022, pp. 1–15. Accessed: Jun. 09, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9843439/