Semi-Autonomous Reconnaissance Satellite: A Laboratory Demonstration of End-to-End Onboard Mission Autonomy

Shtoyerman Eran[1], Gal Pinchas[1], Shterman Doron[1],
[1]IMAGESAT International, 4 Yoni Netanyahu St. Or Yehuda, Israel.

ABSTRACT

Real-time responsiveness to user-driven tasking remains limited in current space-based reconnaissance systems due to constrained communication bandwidth and reliance on ground-based mission planning. This work presents a laboratory demonstration of a semi-autonomous satellite architecture capable of executing high-level, natural-language mission requests entirely onboard. The system introduces a Natural Language Mission Interface (NLMI), enabling compact "chat-like" tasking over low-bandwidth links, which is interpreted by an onboard AI agent that decomposes the request into perception, reasoning, and control tasks.

The processing chain integrates multi-modal data ingestion, AI-based object detection and classification, and closed-loop prediction logic, all implemented on an edge-computing platform representative of future satellite payloads. Experimental results using real satellite imagery demonstrate high detection performance for maritime targets and significant reduction in downlink requirements through transmission of compressed image "stamps" and metadata only.

These results validate the feasibility of closed-loop onboard autonomy for responsive space-based reconnaissance systems.

Keywords: *On-Orbit Processing, Edge AI, Autonomous Satellites, Multimodal Intelligence*

1. INTRODUCTION

Real-time responsiveness to user-driven tasking remains a fundamental limitation in current space-based reconnaissance systems. Such capabilities are largely confined to low-altitude airborne platforms, primarily due to constrained satellite communication bandwidth, intermittent ground contact, and the continued reliance on ground-based mission planning and data processing. As a result, existing satellite architectures are inherently limited in their ability to support time-critical, interactive operations.

Enabling true responsiveness requires a paradigm shift toward highly autonomous space systems, in which sensing, interpretation, decision-making, and control are performed directly onboard the satellite. Recent advances in onboard processing, edge artificial intelligence, and spaceborne surveillance architecture have begun to address elements of this challenge [1] - [3]; however, a fully integrated, closed-loop autonomous capability remains largely unrealized.

Consider, for example, a high-level user request to detect a specific class of maritime target within a defined region, estimate its motion from imagery, continuously track it over time, predict its future location across successive orbital passes, and autonomously re-task the sensor to reacquire the same target. Such complex mission objectives can be expressed as compact, natural-language ("chat-like") commands transmitted over narrow-band communication links, potentially from handheld terminals. Realizing this capability requires the satellite to execute the entire chain of perception, reasoning, prediction, and control autonomously, without reliance on ground intervention.

2. LAB DEMONSTRATION

A laboratory demonstration was conducted to validate an end-to-end processing chain for a semi-autonomous microsatellite capable of executing complex, high-level user requests. The demonstrated system implements all major stages of perception, reasoning, and data handling onboard, as summarized in Fig. 1.

The processing pipeline consists of the following stages (numbers correspond to the flowchart in Figure 1.):

1) Input Acquisition

Input data included:

   a) High-resolution multispectral imagery acquired using ISI's EROS satellite (approximately 30 cm ground sampling distance), comprising four channels: R, G, B, and NIR (Figs. 2 and 3).

   b) Corresponding navigation data derived from GPS and star tracker measurements.

c) Auxiliary data tables containing relevant target characteristics (e.g., vessel dimensions and attributes).

2) User Prompt (Natural Language Tasking)

Mission objectives were defined as concise natural-language commands. For the current demonstration the prompt contained the following instructions: “Find a tanker ship approximately 270 meters long with a helicopter deck. Predict its location during the satellite’s next orbit. Maneuver the cameras at the correct time so that the ship remains within the field of view. If there are no clouds locate the specific ship. Track it for as long as possible. Send compressed image ‘stamps’ of the ship at different locations to the ground station.”

3) Image Preprocessing

Raw imagery was processed using standard radiometric and image enhancement techniques, including non-uniformity correction (NUC), denoising, and normalization, to improve data quality and ensure compatibility with downstream analysis modules.

4) Task Deciphering and Distribution

An onboard AI agent interprets the user prompt and decomposes it into a structured sequence of tasks. Each task is then assigned to the appropriate processing module or AI model. Multiple candidate foundation models are currently under evaluation for this function.

5) Information Extraction

Relevant features are extracted from the imagery using dedicated algorithms and AI models. These include object-level attributes and, when temporal data are available, motion characteristics such as velocity. In the present demonstration, direct velocity estimation from satellite imagery was limited; therefore, motion estimation capabilities were validated using video data acquired from other satellite platforms.

6) Object Detection

Two complementary approaches were evaluated:

a) A dedicated YOLO-based detector [4], trained specifically for maritime target detection, providing high detection accuracy (Figs. 2 and 3).

b) A vision-language model (VLM), enabling more general object detection and semantic interpretation (Fig. 9), evaluated using open-source models such as Qwen-VL [5], Gemma 4 [8] and SAM [6].

In an operational system, the onboard AI agent selects the appropriate model based on mission requirements and scene characteristics.

7) Object Classification

Detected objects were further classified using:

a) A dedicated EfficientNet-based classifier [7], trained for maritime target classification.

b) A VLM-based approach for generalized classification tasks. As in the detection stage, model selection is managed by the onboard AI agent.

8) Prediction Logic

Future target locations were estimated using classical prediction algorithms that combine navigation data with inferred or externally sourced motion information. In this demonstration, vessel velocity was derived from AIS data due to the limited availability of temporal image sequences. Ongoing work explores AI-based approaches for fully image-driven prediction.

9) Camera Maneuvering Logic

Based on predicted target locations and satellite state information, optimal sensor pointing commands were computed using classical control algorithms. This component was not executed on the same edge-computing platform in the current demonstration and is therefore not included in the onboard validation results.

10) Output Processing

Mission-relevant outputs were selected, compressed, and encrypted for transmission over constrained communication links. The output data included geolocation metadata and compressed image “stamps” of detected targets (Figs. 4-8).

11) Data Transmission

The processed data were transmitted over a secure, narrow-bandwidth communication channel, either to ground stations or to other space assets. In the current demonstration, the narrow-band channel was simulated using a dedicated Ethernet connection.

3. LAB SETUP

All processing stages were implemented on an edge-computing platform based on the NVIDIA Jetson Orin system-on-module, representative of the class of high-performance embedded processors targeted for next-generation satellite payload computers. The platform provides GPU-accelerated inference capabilities within a constrained size, weight, and power (SWaP) envelope, enabling real-time execution of multiple AI models alongside classical processing algorithms. The software stack included optimized inference runtimes and parallel processing pipelines to support concurrent execution of perception, reasoning, and data handling tasks.

The input dataset consisted of real multispectral imagery acquired by ISI's very-high-resolution EROS satellites, together with corresponding satellite navigation data derived from onboard GPS and star tracker systems. The imagery includes four spectral channels (R, G, B, and NIR) at approximately 30 cm ground sampling distance, providing sufficient spatial and spectral fidelity for detailed target analysis. In the representative scenario evaluated in this work, two images of nearby geographic area were acquired during consecutive orbital passes, separated by approximately 90 minutes. This configuration enables evaluation of target detection, identification, and re-acquisition across temporal gaps consistent with typical low Earth orbit revisit intervals.

To emulate operational communication constraints, interaction with the edge-computing platform was conducted over a network interface configured to simulate a narrow-bandwidth satellite link with a data rate of 1 Mbit/s. This setup reflects realistic limitations on uplink command transmission and downlink data dissemination and allows assessment of the system's ability to operate under strict bandwidth constraints. In particular, the use of compact natural-language tasking and selective transmission of compressed outputs was validated under these conditions.

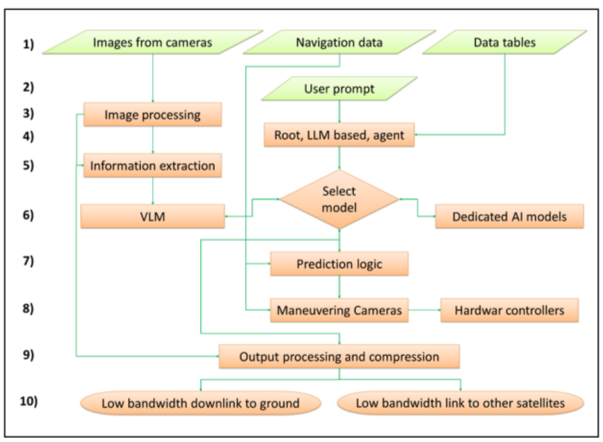

Figure 1: Processing flow chart.

Although the current implementation was conducted in a laboratory environment, the selected hardware, data sources, and communication constraints were chosen to closely approximate operational conditions expected in future small-satellite systems, thereby providing a representative validation of the proposed architecture.

4. RESULTS

Results obtained from the dedicated YOLO-based detection model are presented in Figs. 2 and 3. The model, trained on ISI's maritime dataset, achieved a detection rate of approximately 98% for vessels longer than 100 m under nominal imaging conditions. Performance degradation was observed in the presence of cloud cover and reduced contrast, leading to occasional false detections, as illustrated in Fig. 8.

Target identification was performed by combining geometric and morphological features extracted by the detection stage (e.g., estimated length, width, and aspect ratio) with a dedicated EfficientNet-based classification model. This approach enabled reliable classification of the detected vessel as a tanker in the evaluated scenario, demonstrating the effectiveness of combining specialized detection and classification pipelines.

In addition to dedicated models, several open-source vision-language models (VLMs) were evaluated for generalized detection and identification tasks. These models were not fine-tuned on the ISI dataset. An example of airplane detection using the Qwen-VL model is shown in Fig. 9, where the system successfully interprets a natural-language prompt and identifies multiple objects in the scene.

While the VLMs demonstrated reasonable performance under favorable imaging conditions, their robustness degraded in more challenging scenarios, including low-contrast imagery and cluttered environments. Furthermore, inference latency on the edge-computing platform was observed to be approximately one to two orders of magnitude higher than that of the dedicated YOLO-based detector, limiting their suitability for time-critical onboard applications.

Overall, the results highlight a clear trade-off between model flexibility and operational performance. Dedicated models provide high accuracy and low-latency inference for well-defined tasks, whereas VLMs offer increased adaptability at the cost of reduced efficiency and robustness under constrained onboard conditions.

Fig. 2: Image of see area from Eros satellite, red circle marks the target ship, yellow stars marks the location of other detected ships.

Fig. 3: image of the predicted area of the target ship in the next orbit of the satellite. The target ship is marked in red circle

Fig 4: The target ship in the first orbit

Fig 5: The target ship in the second orbit

Fig 6: Another ship detected in the 2nd image

Fig 7: Another ship detected under clouds

Fig 8: Example of false detection

Fig. 9: Example of object detection using a vision-language model (Qwen-VL) under natural-language prompting

5. DISCUSSION

The presented architecture represents a shift from conventional ground-centric satellite operations toward a closed-loop onboard autonomy paradigm. By executing perception, reasoning, and decision-making directly on the satellite, the system significantly reduces latency between user intent and actionable output. This enables near real-time responsiveness, which is critical for time-sensitive applications such as maritime tracking and disaster response.

A key advantage of the approach is efficient utilization of communication bandwidth. By transmitting only mission-relevant outputs—such as compressed image "stamps" and metadata—the system reduces dependence on high-rate downlink channels, making it well suited for small-satellite platforms with limited communication capacity. In addition, the use of natural-language tasking simplifies user interaction and allows flexible, high-level mission definition without predefined command structures.

These benefits, however, introduce several trade-offs. Onboard AI processing is constrained by available computational resources, power, and thermal limits. While dedicated models provide efficient and reliable performance, more general-purpose models, such as vision-language models, incur significantly higher latency and are less robust under challenging imaging conditions. This highlights the need for hybrid architectures that balance performance and flexibility.

Another important challenge is ensuring the robustness of autonomous decision-making. Errors in detection, classification, or task decomposition may propagate through the closed-loop system, particularly under degraded sensing conditions such as cloud cover or low contrast. Mitigating these risks requires further development of confidence estimation, validation mechanisms, and fallback strategies.

Finally, the current implementation does not fully integrate perception and control on a single platform, as sensor maneuvering was not executed onboard. Achieving full closed-loop autonomy will require tighter coupling between these components, as well as validation under realistic operational constraints, including hardware reliability and environmental conditions.

6. SUMMARY AND FUTURE PLANS

This work demonstrates the feasibility of executing complex, closed-loop reconnaissance tasks entirely onboard a high-resolution satellite platform. The presented system integrates natural-language tasking, AI-driven task decomposition, multi-modal perception, and prediction logic into a unified processing chain operating on an edge-computing architecture representative of next-generation satellite payloads. The results confirm that high-level mission objectives can be translated into autonomous onboard actions, significantly reducing dependence on ground-based planning and post-processing.

Furthermore, the demonstrated capability enables a new class of responsive space systems, in which satellites can interact with users in near real time and adapt dynamically to evolving mission conditions. In the context of the current experiment, this was illustrated through the detection, classification, tracking, and re-acquisition of maritime targets. More broadly, such capabilities have direct applicability to time-critical scenarios including maritime domain awareness, disaster response (e.g., tsunami and flood monitoring), search-and-rescue operations, and homeland security missions, where rapid decision-making and minimal latency are essential. Beyond single-satellite operation, the proposed architecture is particularly well suited for deployment within distributed small-satellite constellations. When combined with inter-satellite communication links, onboard autonomy can enable cooperative sensing, task sharing, and distributed decision-making across multiple space assets, further enhancing coverage, responsiveness, and operational resilience.

While the current study provides a laboratory-scale validation, several aspects require further development. These include full integration of onboard motion estimation from sequential imagery, tighter coupling between perception and control for real-time sensor re-tasking, and improved efficiency and robustness of generalized AI models under constrained onboard conditions. In addition, comprehensive evaluation of system performance under realistic environmental conditions, including varying illumination, cloud cover, and adversarial scenarios, remains an important area for future work.

The next phase of this effort will focus on on-orbit validation using an operational satellite platform. This step will enable end-to-end assessment of the proposed architecture under true space conditions, including real communication constraints, orbital dynamics, and hardware limitations. Successful demonstration in orbit will constitute a critical milestone toward the deployment of operational semi-autonomous reconnaissance capabilities in future ISI satellite systems.

ACKNOWLEDGMENTS

The authors thank the ImageSat International (ISI) engineering team for access to the satellite imagery and payload processing hardware used in this work, and for their support in conducting the onboard inference evaluation campaign.

REFERENCES

[1] B. Liu, X. Zhang, K. Tang, M. Liu, and L. Liu, “Spaceborne video-SAR moving target surveillance system,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), 2016, pp. 2348–2351.

[2] G. Labrèche, D. Evans, D. Marszk, T. Mladenov, V. Shiradhonkar, T. Soto, and V. Zelenevskiy, “OPS-SAT spacecraft autonomy with TensorFlow Lite, unsupervised learning, and online machine learning,” in Proc. IEEE Aerosp. Conf., 2022, pp. 1–17.

[3] M. M. Gost, I. Leyva-Mayorga, A. Pérez-Neira, M. Á. Vázquez, B. Soret, and M. Moretti, “Edge computing and communication for energy-efficient earth surveillance with LEO satellites,” in Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), 2022, pp. 556–561.

[4] P. Jiang et al., “A review of YOLO algorithm developments,” Procedia Comput. Sci., vol. 199, pp. 1066–1073, 2022.

[5] P. Wang et al., “Qwen2-VL: Enhancing vision-language model’s perception of the world at any resolution,” arXiv:2409.12191, 2024.

[6] N. Ravi et al., “SAM 2: Segment anything in images and videos,” arXiv:2408.00714, 2024.

[7] M. Tan and Q. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

[8] Gemma Team et al., “Gemma: Open Models Based on Gemini Research and Technology,” arXiv preprint arXiv:2403.08295, 2024.