Are We Ready to Enable Satellite Autonomy Through On-Board Image Registration?

CLAUDIO DI SALVO, Politecnico di Milano, Italy
EMANUELE DEL SOZZO, Massachusetts Institute of Technology, USA
GIUSEPPE SORRENTINO, Politecnico di Milano, Italy
ELEONORA D’ARNESE, University of Edinburgh, United Kingdom
PAOLO PANICUCCI, Politecnico di Milano, Italy
DAVIDE CONFICCONI, Politecnico di Milano, Italy

Enabling autonomous CubeSats requires shifting image-based decision-making from ground stations directly to on-board Commercial-Off-The-Shelf (COTS) hardware to overcome the unsustainable downlink bottlenecks typical of Low Earth Orbit (LEO) missions. A critical yet largely neglected aspect of this transition is Image Registration (IR), which is central for multi-temporal analysis, change detection, and multimodal data fusion. Despite its pivotal role, existing IR methods are designed and evaluated on unconstrained computing platforms, making their readiness for on-board autonomous operations unclear due to the lack of the assessment of latency-accuracy-energy trade-offs. In this paper, we show that state-of-the-art “out-of-the-box” IR algorithms fail to reconcile the latency, accuracy, and energy constraints inherent to COTS-based satellite platforms. We present *STAR-Bench*, an open-source framework for systematic and reproducible characterization of IR methods on on-board hardware across diverse sensing modalities, terrains, and deformation regimes. Our characterization reveals a fundamental specialization gap: lightweight methods meet latency constraints but fail under multimodal conditions; while, robust methods incur up to an order-of-magnitude higher energy costs, precluding effective autonomy. To capture system-level trade-offs, we introduce Quality-Adjusted Cost (QAC), a metric to quantify the energy required to achieve a target accuracy level. Guided by this analysis, we propose *ETNA*, a mission-aware framework to dynamically orchestrate heterogeneous IR pipelines. To address the challenging multimodal scenario, we perform a hardware-software co-design of an FPGA-accelerated solution. *ETNA* reduces latency by up to 1.7× in Synthetic Aperture Radar (SAR)-Optical scenarios and improves QAC by up to 12.03× over all Pareto-optimal baselines, while also improving registration accuracy. Our results demonstrate that satellite autonomy cannot be achieved by deploying existing IR algorithms unchanged; instead, IR must be explicitly co-designed with COTS hardware and mission constraints.

CCS Concepts: • **Applied computing** → **Earth and atmospheric sciences**; • **Computing methodologies** → *Computer vision*; **Image manipulation**; • **Computer systems organization** → **Embedded and cyber-physical systems**; **Parallel architectures**.

Additional Key Words and Phrases: On-board Image Registration, Earth Observation, Benchmarking and Characterization, Hardware-Software Co-design

---

Authors’ Contact Information: Claudio Di Salvo, claudio.disalvo@polimi.it, Politecnico di Milano, Milano, Italy; Emanuele Del Sozzo, Massachusetts Institute of Technology, Cambridge, USA, delsozzo@mit.edu; Giuseppe Sorrentino, Politecnico di Milano, Milano, Italy, giuseppe.sorrentino@polimi.it; Eleonora D’Arnese, University of Edinburgh, Edinburgh, United Kingdom, eleonora.darnese@ed.ac.uk; Paolo Panicucci, Politecnico di Milano, Milano, Italy, paolo.panicucci@polimi.it; Davide Conficconi, Politecnico di Milano, Milano, Italy, davide.conficconi@polimi.it.

ACM 2476-1249/2026/6-ART30
https://doi.org/10.1145/3805628

ACM Reference Format:
Claudio Di Salvo, Emanuele Del Sozzo, Giuseppe Sorrentino, Eleonora D'Arnese, Paolo Panicucci, and Davide Conficconi. 2026. Are We Ready to Enable Satellite Autonomy Through On-Board Image Registration?. *Proc. ACM Meas. Anal. Comput. Syst.* 10, 2, Article 30 (June 2026), 41 pages. https://doi.org/10.1145/3805628

1 Introduction

Earth Observation (EO) is the process of gathering information from space to study our planet and assess the effects of human activities. Over the years, novel small satellites have revolutionized the EO paradigm, collecting vast amounts of images for time-critical applications, such as precision agriculture, disaster relief, and climate change monitoring [5, 7, 10, 15, 20, 22, 38, 46, 61]. In such a context, the proliferation of Commercial-Off-The-Shelf (COTS) components is enabling a new era of low-cost satellites, such as CubeSats [29, 49, 56], which are transforming EO paradigms and enabling fleets of satellites to acquire massive amounts of images [8, 57, 59]. However, satellites in LEO can downlink images during brief and infrequent transfer windows [15, 54, 62], causing a critical **downlink bottleneck** since the data gathered far surpasses the capacity to transmit it [20, 54] Moreover, only a few pixels in the image typically carry valuable information for downlinking, due, for instance, to cloud coverage [27]. Several attempts have been made to overcome the critical bottleneck (e.g., Distributed Downlink [32, 54]), but none of them were able to reduce the information to be downlinked as the Orbital Edge Computing (OEC) [16, 17]. This paradigm rethinks satellites as autonomous agents capable of processing data directly in orbit and transmitting only relevant information, thereby removing ground-processing delays, which become unsustainable for time-sensitive events, such as wildfire detection, oil spill detection, and flood detection [10, 51].

Nevertheless, modern OEC applications, such as real-time disaster monitoring and multi-sensor data fusion [15, 20, 31, 38], rely on a strong assumption: images must be perfectly aligned before any further processing [38]. However, acquisition distortions, sensor drift, or geometric misalignment due to spacecraft motion violate this assumption, as exemplified in Figure 1 and highlighted recently by European Space Agency (ESA) [38, 43]. Therefore, OEC applications require a critical preprocessing step, called Image Registration (IR). Traditionally, IR occurs at ground stations with unlimited resources, but relying on ground-based IR inherently sacrifices spacecraft autonomy and introduces unacceptable delays. Therefore, to enable OEC, IR must execute directly on-board at every acquisition, managing a crucial trade-off between latency, accuracy, and energy consumption.

Despite its crucial importance, on-board IR is a uniquely challenging task because multiple conflicting constraints must be satisfied simultaneously. On-board processing is constrained by Size, Weight, and Power (SWaP) requirements [12], where computational efficiency must be integrated into systems with a small form factor, leading to the preferred use of COTS Field Programmable Gate Arrays (FPGAs). Additionally, IR must be accurate and respect image acquisition latency constraints, ensuring that each group of images is processed before the next group is acquired. Literature highlights 1 to 30 seconds as a reasonable latency constraint for the whole image processing pipeline for OEC [15], possibly leaving for preprocessing tasks such as IR a small margin. Finally, IR is highly sensitive to acquisition sensors; aligning images acquired with the same sensor type (*monomodal IR*), different sensor types (*multimodal IR*), different spectral bands, or different polarity radar acquisition requires distinct strategies [9, 63, 65]. Due to these constraints, there is no unique solution to tackle monomodal and multimodal on-board IR, leading to our core research question:

> Considering the plethora of existing state-of-the-art methods for Image Registration (IR) and the constraints of embedded processing, how far is the IR landscape from leveraging COTS hardware for on-board image alignment?

Current literature suffers from a dichotomy: algorithmic studies optimize for accuracy and performance on high-end GPUs, ignoring on-board energy constraints [38]. In contrast, hardware-centric

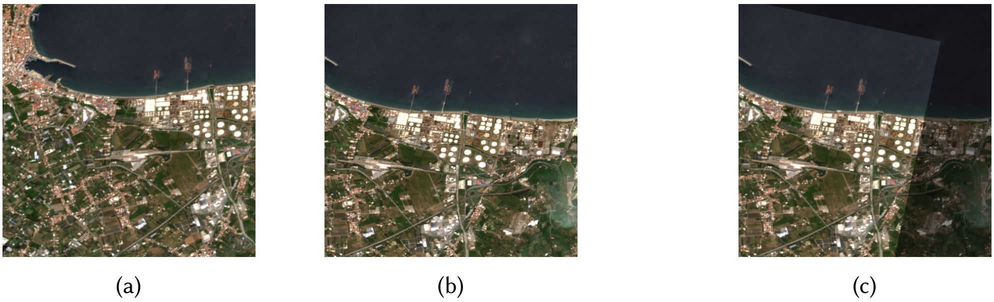


Fig. 1. Visual example of satellite image distortion. (a) The floating image to be registered. (b) The fixed reference image. (c) The fused image shows the successful alignment of the floating image with the reference geometry, strictly required for in orbit operation.

studies focus on FPGA energy efficiency and throughput rather than the registration quality [21]. No existing framework and metrics bridge these two worlds to quantify the critical latency-accuracy-energy trade-off for autonomous missions. To address this research question, we propose a *Satellite Testbed for Alignment and Registration Benchmarking* (*STAR-Bench*) to standardize the quality assessment of existing implementations. *STAR-Bench* enables a fair assessment aligned with stringent orbital constraints by evaluating algorithms across diverse sensing modalities, terrain types, and deformation scenarios on representative space computing COTS hardware [12], such as the Kria KR260 (System-On-Chip (SoC) with FPGA) and the Jetson Xavier NX (Graphics Processing Unit (GPU)). Central to this framework is the Quality-Adjusted Cost (QAC), a novel system-level metric that we propose to model the latency-accuracy-energy trade-off space, addressing a clear gap in the current literature. Unlike scalar and literature [26] metrics, the QAC quantifies the multidimensional frontier of registration tasks, providing a principled method for selecting the most appropriate approaches for OEC pipelines tailored to specific mission profiles. Our characterization (§4) reveals a fundamental *"specialization trade-off"*: feature-based methods minimize latency but suffer functional collapse in multimodal scenarios, whereas robust Deep Learning models (e.g., Transformers) incur prohibitive energy costs that violate CubeSat power budgets. Building on this "specialization trade-off," we propose *ETNA* a framework to orchestrate IR pipelines in the mission context. Within this framework, we implement a hardware-software co-designed execution model to push the boundaries of the latency-accuracy-energy frontier. Our solution provides a low-overhead execution path for monomodal tasks while deploying a high-performance, multimodal pipeline. This is achieved by significantly extending the intensity-based co-design of [13] through a multi-scale pyramidal paradigm [52], paired with a custom-tailored FPGA accelerator specifically redesigned to offload the bottleneck kernels of the pyramidal search.

To summarize, our main contributions are:

- ***STAR-Bench*** (§3): a modular, open-source testbed for the systematic characterization of on-board IR. It introduces the QAC metric, a novel performance indicator that models the latency-accuracy-energy trade-off, enabling the identification of the Pareto-optimal frontier for resource-constrained registration across disparate COTS platforms.
- ***ETNA*** (§5): a heterogeneous framework for constructing and orchestrating adaptive IR pipelines. By introducing a modular abstraction layer, *ETNA* enables the seamless integration of intensity-based and deep learning methods, allowing the system to dynamically adapt its complexity to the requirements of the specific sensing modality.

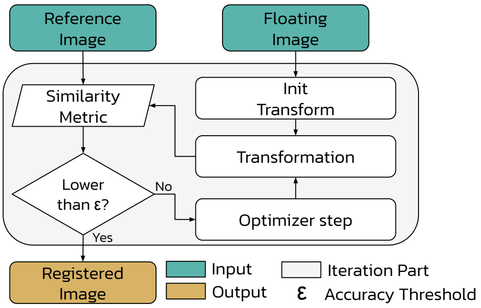


Fig. 2. Intensity-based algorithms iterative procedure. It repeatedly adjusts transformation parameters to maximize a global similarity metric.

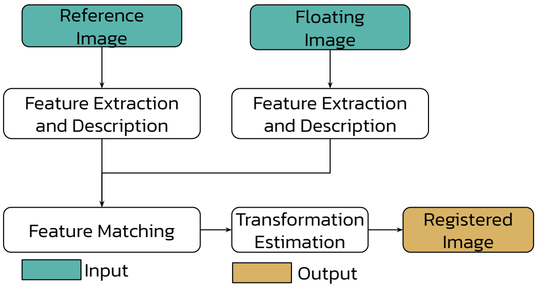


Fig. 3. Standard multi-stage pipeline for feature-based registration. Accuracy depends on the quality of the detected local features.

- ***Co-designed multimodal accelerator*** (§5.2): a hardware-software co-designed solution that pairs a hierarchical software control logic with a custom FPGA acceleration layer. This synergy minimizes redundant computations while maintaining high precision.

ETNA reduces latency by up to 1.7× in Synthetic Aperture Radar (SAR)-Optical scenarios and improves QAC by up to 12.03× over all Pareto-optimal baselines, while also improving registration accuracy. Our results prove that deploying existing IR algorithms is insufficient to achieve satellite autonomy; instead, IR must be explicitly co-designed with COTS hardware and mission constraints.

2 Motivation

2.1 Image Registration Background

Image Registration (IR) is defined as the process of geometrically aligning two or more images of the same scene in a common global reference frame [9, 66]. The objective is to ascertain the optimal set of parameters for a transformation function, T, that maps the coordinate system of each point x_i in the floating image, I_F, to one of the reference image, I_R.

$$I_R(\mathbf{x_i}) \approx I_F(T^*(\mathbf{x_i})) \tag{1}$$

Once T^* is found, the images share a common frame of reference, **which is the essential prerequisite** for all on-board analyses [38, 63–65].

The computer vision field distinguishes three prominent families of algorithms for finding the optimal transformation T^*, depending on the employed methodology.

Intensity-based Methods - they follow a heuristic-based approach to iteratively explore the space of parameters up to convergence, as shown in Figure 2. At each step, the algorithm uses an optimizer to compute candidate registration parameters, applies them through a pixel-wise image transformation, and evaluates the result through a similarity metric [9]. The iterative process targets a specific accuracy threshold, ϵ, which is a hyperparameter that depends on the use case. Thus, this class of algorithms favours high accuracy and robustness at the cost of multiple pixel-wise compute-intensive steps [55].

Feature-based Methods - they assume that images can be aligned by extracting and matching a sparse set of salient features [9, 66]. These approaches typically follow a two-stage pipeline: first, keypoints are detected and encoded into invariant feature descriptors (e.g., SIFT [36]); second, these descriptors are matched to establish candidate correspondences. To ensure geometric consistency, outlier rejection schemes such as RANSAC [25] are employed to filter erroneous matches. Unlike intensity-based methods that require exhaustive pixel-wise optimization, feature-based approaches leverage sparse representations for greater computational efficiency. However, their success depends on the presence of repeatable textures in both images.

Deep Learning-based Methods - they use Deep Neural Networks (DNNs) to learn the registration process from large-scale datasets [4, 40]. State-of-the-art approaches aim to surpass traditional pipelines by extracting semantic feature descriptors that are robust to geometric and radiometric distortions [37]. However, these data-driven models inherently struggle with generalization [18]. Despite their superior performance on monomodal tasks, their applicability to multimodal scenarios is often limited by the lack of aligned ground-truth data and significant domain shifts [65].

2.2 The On-Board Computing Challenge

Traditionally, IR is performed on ground stations, where the absence of strict SWaP constraints allows for the use of complex, computationally intensive algorithms to maximize alignment precision. However, this ground-based approach needs the costly, slow downlink of raw data that prevents from moving all the acquired imagery. Consequently, less than 2% of acquired data can be transmitted [15]. This bottleneck requires shifting the IR pipeline directly on-board. Nevertheless, deploying IR algorithms on small form-factor satellites, such as CubeSats, is dictated by strict SWaP budget [12]. While Size and Weight are static physical constraints, **power is the critical dynamic variable** governing compute capacity. With power budgets often constrained to $< 10W$ [12], standard high-performance accelerators are infeasible. Furthermore, LEO satellites maintain high orbital velocities ($> 7km/s$ [3]), needing rapid processing. The image processing must complete before the next acquisition, enforcing a strict 1 to 30 seconds deadline [15]. Finally, as registration errors propagate through the image processing pipeline, accuracy remains paramount. Consequently, designing for OEC moves beyond single-objective optimization; solutions must be evaluated for Pareto optimality in the multidimensional latency-accuracy-energy trade-off.

3 *STAR-Bench*: A Framework for Rigorous On-Board Characterization

To standardize the evaluation landscape across disparate and diverse performance metrics and COTS hardware, we introduce Satellite Testbed for Alignment and Registration Benchmarking (***STAR-Bench***), an open-source framework designed for the holistic and rigorous characterization of literature IR algorithms[1]. The design of *STAR-Bench* is grounded in three core principles: completeness, extensibility, and reproducibility. To ensure a comprehensive characterization, we evaluate the proposed solutions across a three-tier taxonomy of operational scenarios: (I) Nominal, (II) Off-Nominal, and (III) Anomaly. As detailed in §3.1, this approach enables rigorous stress testing of registration resilience under increasingly adverse acquisition conditions. Complementing this, *STAR-Bench* ensures extensibility through a modular architecture that facilitates the seamless integration of novel algorithms alongside established baselines. Finally, by providing a unified testbed with pre-packaged, versioned datasets, *STAR-Bench* guarantees reproducibility, ensuring that all reported metrics can be strictly verified on COTS hardware.

We implement *STAR-Bench* in Python, structuring it as a set of three modules that abstract the complexity on the user level. Figure 4 highlights *STAR-Bench* structure:

- **Data Augmentation** Ⓐ: Responsible for generating synthetic test cases by using geometric transformations (§3.1).
- **Algorithm Benchmarking** Ⓑ: It manages experiments by loading the deformed images produced by Ⓐ. To ensure consistency, it enforces a unified interface for all methods, captures outputs, and forwards them to the analysis module Ⓒ (§3.2).

---

[1] Our framework is available at: https://github.com/necst/STAR-Bench

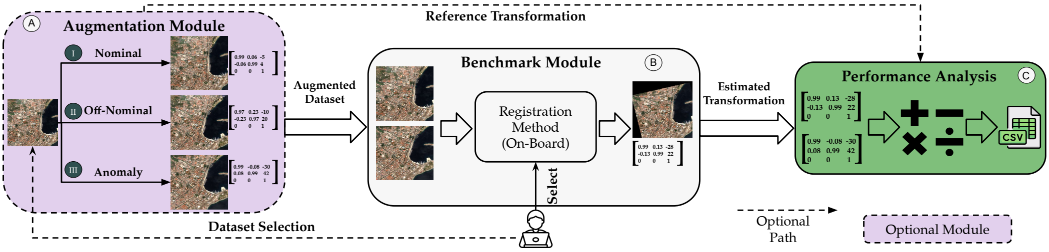


Fig. 4. *STAR-Bench* modular structure. It features three distinct modules to benchmark on-board registration algorithms. The Data Augmentation Module Ⓐ performs geometric transformations on source datasets to simulate operational scenarios. The Benchmarking Module Ⓑ takes as input the dataset and executes the target algorithms directly on the chosen hardware. Finally, the Performance Analysis module Ⓒ, which evaluates the resulting alignments against the generated ground truth of accuracy and efficiency metrics.

- **Performance Analysis** Ⓒ: Computes the suite of metrics. It also supports user-defined metrics, enabling the evaluation of custom Key Performance Indicators (KPIs) while maintaining a standardized testing environment (§3.3).

This modular architecture enables isolated testing and the addition of new components without altering the core evaluation logic, thereby providing a stable and extensible platform for future research and the whole community.

3.1 Data Augmentation

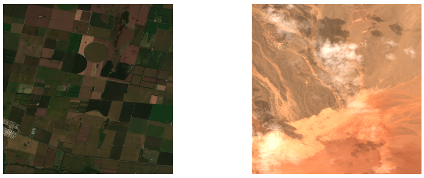
STAR-Bench data augmentation module Ⓐ is an optional module that can be used to simulate different operational scenarios. It starts from an input dataset and applies different transformations to rigorously test algorithm performance across the full spectrum of on-board challenges (§2.2). We articulate such challenges along three axes: terrain diversity, deformation complexity, and spectral, modal, and polarimetric diversity. To base our evaluation on real-world data, we used images acquired through the **Sentinel-1 and Sentinel-2 satellites** [1, 19, 28, 48].

Terrain Diversity - Algorithm performance is strictly conditioned by the spatial and spectral characteristics of the underlying terrain. To ensure generalization beyond idealized benchmarks, our dataset spans four distinct land-use classes (Figure 5) distributed across diverse geographical zones[2]. **Urban environments** provide clear geometric features (e.g., road, roundabout). Conversely, feature-poor environments dominate the remaining classes: **Agricultural** and **Grassland** exhibit high homogeneity and are prone to extreme seasonal variance (e.g., crop cycles). In contrast, **Barren Land**, although texture-deprived, offers highly stable geological formations that are useful for long-term landmark tracking.

Deformation Complexity To simplify the complex 3D modeling involved in image acquisition, 2D-2D homographies are used to simulate how images would be acquired from different positions and orientations. Despite introducing interpolation errors, this approach enables the rapid generation of multiple test cases without requiring complex rendering tools. *STAR-Bench* defines three distinct **operational scenarios** to evaluate algorithm resilience. The primary class, designated as the Ⅰ **Nominal** scenario, reflects standard conditions with perturbations limited to translations of $[-10, 10]$ pixels and rotations of $[-5°, 5°]$. To evaluate resilience against significant attitude control drift, the Ⅱ **Off-Nominal** scenarios expand the search space to translations of $[-20, 20]$ pixels and $[-10°, 10°]$. Finally, the Ⅲ **Anomaly** tier simulates critical physical failure modes, such as

---

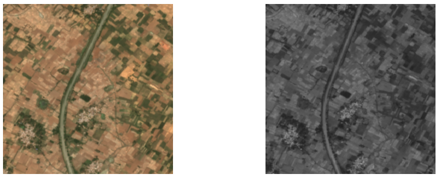
[2] A map and a table of the geographical distribution are provided in the appendix Appendix A.1



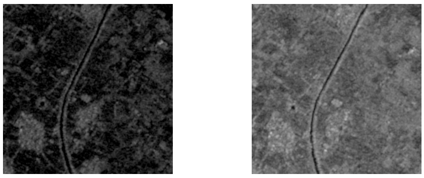
(a) Agricultural ⓐ (b) Barrenland ⓑ

(c) Grassland ⓒ (d) Urban ⓓ

Fig. 5. Examples of terrain categories from the dataset: (a) Agricultural, (b) Barrenland, (c) Grassland, and (d) Urban.



(a) RGB ①, ③ (b) Band B3 ②



(c) SAR-VH ①, ③, ④ (d) SAR-VV ①, ③, ④

Fig. 6. Dataset imaging modalities and registration scenarios. Markers indicate the participation of each data type in the evaluated tasks, as defined in §3.1.

severe attitude control drift, by applying extreme deviations within $[-50, 50]$ pixels and $[-15°, 15°]$. Within each scenario, transformation parameters are generated via Monte Carlo simulation. This approach ensures a thorough exploration of the algorithmic search space, allowing our framework to probe genuine operational corner cases rather than only nominal acquisition conditions.

Acquisition Diversity - To facilitate a direct mapping between image modalities and the evaluated registration tasks, we introduce a taxonomy based on symbolic markers (① – ④), as illustrated in Figure 6. Specifically, we categorize the input data based on their role in four registration scenarios: ① **monomodal** registration, which aligns temporal sequences from the same sensor to assess resilience against temporal decorrelation and seasonal variations; ② **cross-band** alignment of distinct spectral channels (e.g., Visible vs. NIR) to evaluate the handling of gradient inversions; ③ **multimodal** Registration, targeting Optical-to-SAR alignment by addressing the non-linear radiometric shifts between passive and active sensing; and ④ **multipolarization**, which leverages SAR Vertical-Vertical (VV) and Vertical-Horizontal (VH) channels to test against surface scattering from artificial structures (VV) and volumetric scattering from vegetation (VH) [53]. Combined with the deformation tiers, *STAR-Bench* yields ≈24,000 256×256 image pairs [45], ensuring statistically significant characterization across all orbital conditions.

3.2 Algorithm Benchmarking

Differences in input pre-processing, pixel representation, or data layout can bias the evaluation of registration algorithms. To mitigate these artifacts, the *STAR-Bench* benchmarking orchestrator Ⓑ enforces a unified I/O interface. All synthetic test cases generated by Ⓐ can be loaded via a strict wrapper that standardizes the tensor format[3] before dispatching them to the registration kernels. This architecture ensures that performance variances are attributable to the registration logic.

Furthermore, while recent literature emphasizes the role of pre-processing in accelerating IR [30, 50], *STAR-Bench* explicitly decouples data enhancement from registration. The framework allows users to inject custom pre-processing blocks (e.g., denoising, super-resolution) into the pipeline in a controlled manner. However, to establish a rigorous baseline for this study, we deliberately bypass

---

[3] The base class implementation and standardization logic are detailed in appendix A.2.

optional enhancement steps and evaluate the “**out-of-the-box**” performance of the literature solutions in their native state.

3.3 Performance Analysis

The performance analysis module Ⓒ uses various metrics to assess the quality of the registration pipelines benchmarked in Ⓑ §3.2. The module is designed to map results against an operational viability region, which represents the performance envelope defined by the latency and accuracy constraints of the on-board platform.

We define **latency** as the **end-to-end execution time** required to estimate the registration parameters from input data. To ensure measurement fidelity and isolate the algorithmic computational overhead from storage I/O bottlenecks, all input tensors are preloaded into the system memory before execution. We adopt latency as the primary constraint defining the operational viability region. Specifically, we establish a threshold of 15s, which represents 50% of the 30s upper bound typically allocated for end-to-end on-board image processing pipelines [15].

To assess the quality of our solutions, we consider the **Target Registration Error (TRE) accuracy metric,** computed as in Equation (2).

$$TRE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_p \cdot x_i - T_g \cdot x_i)^2} \tag{2}$$

Such a metric considers a set of N ground-truth correspondence points that are synthetically determined during the data augmentation step Ⓐ. The idea is to measure the Euclidean distance between the ground-truth transformed points ($T_g \cdot x_i$) and the algorithm’s predicted transformed points ($T_p \cdot x_i$). In monomodal, intra-band, and multipolarization scenarios, we assume a Ground Sample Distance (GSD) of 10m, following the spatial resolution of Sentinel-2 multispectral bands. For ③ multimodal SAR-Optical registration, we reconcile the disparate sensor resolutions by mapping all data onto a common 10m. Given the lack of established benchmarks for on-board multimodal (③), intra-band (①), and multipolarization (④) registration, we define our operational targets through an empirical performance envelope derived from the current state-of-the-art ceiling. This baseline identifies the best-case accuracy achievable by existing algorithms, which we use to delineate the operational viability region under our 15s latency constraint. Conversely, for the monomodal scenario (②), we adhere to standardized literature benchmarks [41] by adopting a 2px accuracy threshold. In the context of our evaluation framework, this requirement corresponds to a maximum ground displacement of 20m, providing a stringent metric for assessing whether IR pipelines can satisfy the mission-critical precision requirements for autonomous EO operations.

Since our system targets on-board processing with a minimal power budget, we consider both **power and energy consumption**, $\bar{E} = \bar{P} \cdot Latency$. Ⓒ relies on the target device vendor utility to quantify the power consumption. Furthermore, to address potential resource contention during the execution of concurrent on-board tasks, a detailed analysis can be found in Appendix B. For the AMD Kria KR260, the provided utility offers real-time access to the on-chip power rail sensors[4]. For the Nvidia Jetson Xavier NX, we used the open-source tool Jetson Stats[5]. However, the Ⓒ module can be extended and adapted to different platforms by using the corresponding vendor utility. While effective in general-purpose computing, traditional performance indicators fail to capture the multidimensional trade-off constraints of OEC. To bridge this gap, we propose the Quality-Adjusted Cost (QAC), a system-level metric that model the multidimensional latency-accuracy-energy trade-off (Equation (3)). By treating energy and TRE as equivalent operational penalties, QAC enable a

---

[4] https://github.com/Xilinx/xlnx_platformstats

[5] https://github.com/rbonghi/jetson_stats

principled comparison across disparate hardware architectures. For instance, it allows researchers to distinguish between heuristics that prioritize throughput and energy-intensive models optimized for high-precision ceilings. In this formulation, a higher QAC value indicates a more favorable latency-accuracy-energy profile.

$$QAC = \frac{c}{\bar{E} + \text{TRE}} \tag{3}$$

To ensure numerical stability and facilitate comparison of results across diverse hardware platforms, we use a constant scaling factor, $c = 100$, for all reported results. Collectively, these metrics allow *STAR-Bench* to characterize the Pareto-optimality of registration pipelines and identify the most suitable candidate for specific mission-profile constraints.

4 Characterization of the State of the Art

We leverage *STAR-Bench* to conduct a systematic evaluation of representative state-of-the-art IR pipelines, as detailed in §2.1, under operationally realistic conditions. Rather than proposing a new standalone algorithm, our objective is to stress-test widely adopted, open-source implementations under operationally realistic constraints. We aim to determine whether these **out-of-the-box** solutions can satisfy the latency-accuracy-energy trade-off that define the EO missions (§2.2). To characterize the performance landscape of on-board IR, we evaluate all methods across two heterogeneous COTS platforms: the AMD KRIA KR260 and the NVIDIA Jetson Xavier NX. We establish a common baseline for software-only execution by profiling all non-accelerated implementations on the Kria ARM Cortex CPU. Simultaneously, we evaluate accelerated versions of the pipelines by offloading compute-intensive kernels to specific hardware targets: an FPGA on the Kria and a GPU on the Jetson. This benchmarking approach allows us to decouple the inherent efficiency of the algorithms from the benefits of architectural specialization, exposing the non-linear latency-accuracy-energy trade-off across distinct COTS hardware. We evaluated three primary IR paradigms:

Intensity-Based Methods - We evaluate Faber [13] as a representative state-of-the-art FPGA accelerator for the intensity-based paradigm. By offloading compute-intensive similarity metrics, specifically Cross Correlation (CC) and Mutual Information (MI), to a dedicated FPGA fabric, Faber enables the efficient execution of heuristic optimization strategies. This approach overcomes the latency bottlenecks of the Powell's method, which is traditionally prohibitive on embedded CPUs.

Feature-Based Methods - We evaluate RIFT2 [33] as the state of the art for multimodal robustness based on phase congruency. Since the original reference implementation of the benchmark was restricted to `MATLAB`, preventing direct deployment and profiling on our resource-constrained hardware, we developed a high-fidelity Python port. This port, validated for functional equivalence, enables the native execution of the *STAR-Bench* pipelines on the embedded platforms. To establish a baseline for computational overhead versus robustness, we additionally benchmark other methods such as SIFT [36], SURF [6], and ORB [47].

Learning-Based Methods - We evaluate XFeat [42] as the efficiency-oriented baseline. In contrast, SuperPoint with LightGlue (SP+LG) [35] represents the state of the art in sparse matching precision and has been adopted by the ESA as reference in image alignment [38]. To specifically tackle the multimodal scenario, we integrate MINIMA [44] variants. Unlike other models, MINIMA is expressly designed and fine-tuned for multimodal image registration. We evaluate three configurations ($MINIMA_{LG}$, $MINIMA_{LoFTR}$, and $MINIMA_{RoMa}$); for the last, we employ the "tiny" configuration [23, 24]; notably, its memory footprint exceeds the physical capacity of the Jetson Xavier NX, identifying a critical deployment bottleneck for dense matching models. We specify that $MINIMA_{RoMa}$ (tiny) builds on top of a re-trained multimodal XFeat [23], thus allowing us to

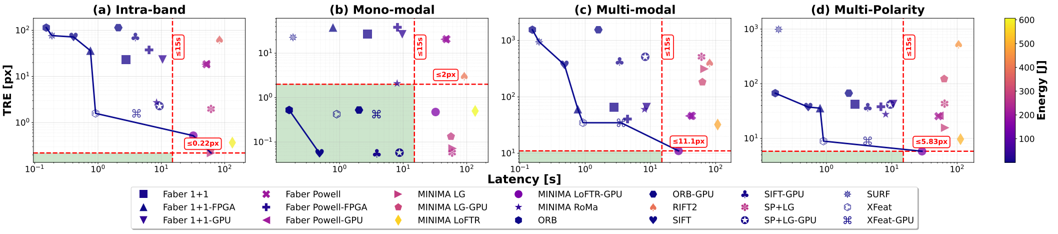


Fig. 7. Pareto characterization of the latency-accuracy-energy trade-off: (a) Intra-band ①, (b) Monomodal ②, (c) Multimodal ③, and (d) Multi-polarity ④ scenarios. Accuracy is the mean TRE across deformations; marker colors indicate the average energy consumption per task [J], while marker color in the legend is the average among the scenarios. The red dashed lines delineate the operational viability region, bounded by a ≤ 15s latency constraint and modality-specific accuracy thresholds (ranging from ≤ 0.22px to ≤ 11.1px). The vertical migration of the Pareto frontier from (a) to (c) quantifies the "specialization trade-off" induced by the radiometric domain gap between optical and SAR sensors. Table 1 reports the Pareto frontier measurements.

consider both monomodal and multimodal XFeat versions. All models are evaluated in a zero-shot setting to rigorously test their generalization capabilities under real-world orbital-domain shifts.

Evaluation Strategies: we evaluate the complete IR pipelines across three orthogonal dimensions: sensing modalities (①-④, §3.1); geometric resilience (I-III, §3.1); and environmental diversity, (ⓐ-ⓓ, §3.1). This structured approach allows *STAR-Bench* to systematically explore the exhaustive combinatorial space generated via the data augmentation engine (§3.1). Performance is quantified through the individual metrics of TRE, latency, and energy, while their joint trade-offs are synthesized via the proposed QAC. The system characterization is reported by averaging across all image pairs. Detailed analysis for all benchmarked modalities and scenarios are in Appendix B.

4.1 *STAR-Bench*- Analysis Across Sensing Modalities

Table 1. Comparative analysis of Pareto optimal method across sensing modalities. ① Intra-band, ② Monomodal, ③ Multimodal, ④ Multi-Polarity. Best results in **bold**.

Type	Pipeline	Lat. [s]	TRE [px]	En. [J]	QAC
①	ORB	**0.15 ± 0.07**	114.1 ± 1384.6	**0.37 ± 0.24**	0.87
	SURF	0.19 ± 0.09	76.7 ± 1160.3	0.54 ± 0.34	1.30
	SIFT	0.41 ± 0.18	71.7 ± 909.3	1.44 ± 0.68	1.37
	Faber 1+1-FPGA	0.76 ± 0.05	36.2 ± 30.9	2.48 ± 0.14	2.59
	XFeat	0.92 ± 0.11	1.59 ± 10.00	2.15 ± 0.41	**26.7**
	MINIMA LoFTR-GPU	31.9 ± 9.30	0.52 ± 1.82	157.2 ± 42.5	0.63
	MINIMA LG	60.9 ± 2.99	**0.22 ± 0.49**	267.7 ± 26.7	0.37
②	ORB	**0.16 ± 0.07**	0.53 ± 5.72	**0.38 ± 0.25**	**109.9**
	SIFT	0.48 ± 0.21	**0.06 ± 0.13**	1.70 ± 0.82	56.8
③	ORB	**0.16 ± 0.06**	1548.4 ± 6383.7	**0.50 ± 0.21**	0.06
	SURF	0.20 ± 0.09	947.3 ± 1797.8	0.72 ± 0.35	0.11
	SIFT	0.48 ± 0.19	369.8 ± 433.7	1.92 ± 0.77	0.27
	Faber 1+1-FPGA	0.77 ± 0.18	60.6 ± 59.4	2.56 ± 0.87	1.58
	XFeat	0.93 ± 0.12	34.9 ± 32.5	2.18 ± 0.47	**2.70**
	XFeat-GPU	3.56 ± 3.80	34.8 ± 32.5	2.77 ± 6.11	2.66
	MINIMA LoFTR-GPU	27.0 ± 7.35	**11.1 ± 22.9**	133.3 ± 32.7	0.69
④	ORB	**0.16 ± 0.06**	67.2 ± 331.8	**0.47 ± 0.17**	1.48
	SIFT	0.53 ± 0.22	37.4 ± 385.1	1.93 ± 0.74	2.54
	Faber 1+1-FPGA	0.80 ± 0.28	35.8 ± 30.5	2.68 ± 1.39	2.60
	XFeat	0.90 ± 0.09	8.92 ± 24.8	2.12 ± 0.28	**9.06**
	MINIMA LoFTR-GPU	29.0 ± 8.53	**5.83 ± 16.9**	143.2 ± 38.2	0.67

This section characterizes the performance of the selected pipelines across the four sensing modalities (①-④) by averaging results across all geometric deformation (I-III) tiers to establish a baseline for algorithmic deformation robustness. The interaction between registration accuracy (TRE), latency, and average energy expenditure per task (represented by the marker color gradient) is illustrated in Figure 7 and in Table 1.

Intra-band Registration (①). Serves as the baseline for evaluating system performance under minimal information differences, yet the Pareto distribution reveals a significant gap between efficiency and reliability. While MINIMA LG achieves an absolute accuracy ceiling of 0.22px, it incurs a prohibitive computational cost of 60.86s and 267.70J per task. Conversely, the latency floor established by ORB at 0.15s comes at the expense of instability, yielding a TRE of 114.13px. In this gap, the solution that stands out is XFeat, which, with 0.92s and 1.59px, represents a Pareto-optimal point within the latency-accuracy-energy trade-off, as it minimizes energy-per-registration while maintaining a competitive accuracy profile.

Monomodal Registration (②). Evaluates temporal sequences captured by the same sensor, a scenario where a significant number of benchmarked methods successfully converge within the 2px accuracy requirement. In this domain, a clear architectural divide emerges, with feature-based methods defining the Pareto frontier and outperforming deep learning-based benchmarks in both execution speed and energy efficiency. ORB establishes the latency floor for this modality at 0.16s with a mean TRE of 0.53px. Transitioning from ORB to SIFT yields a substantial 8.83× accuracy improvement, achieving near-optimal 0.06px accuracy, at the cost of a 3× increase in latency (0.48s) and an energy expenditure of 1.70J. These results prove that for high-texture monomodal data, classical descriptors remain the most effective solution for resource-constrained hardware. More sophisticated models, such as $SP + LG$, only enter the operational viability region (defined by the red dashed lines in fig. 7(b)) when leveraging GPU accelerated.

Multimodal Registration(③). Represents the primary "specialization trade-off" in our study, where the inherent domain gap between optical and SAR sensors leads to complete registration failure for certain methods. As shown in the Pareto distribution, feature-based methods like ORB and SIFT fail to register, exhibiting mean errors of 1548.41px and 369.78px, respectively. The first significant reliability breakthrough occurs with the Faber 1+1-FPGA module, which achieves a 6.10× accuracy improvement over SIFT, reducing the error to 60.57px in only 0.77s. The transition to a deep-learning method further improves the TRE, as XFeat with a mean TRE of 34.85px. While MINIMA LoFTR-GPU reaches the absolute accuracy ceiling of 11.06px, its prohibitive latency of 26.98s and energy cost of 133.33J make it unsuitable for real-time OEC tasks. Thus, accuracy remains inherently capped by the larger GSD, which limits discernible alignment details.

Multipolarization Registration (④). Evaluates the system's ability to reconcile images captured by the same sensor but across different polarization channels, which introduces significant scattering asymmetries. As shown in the multimodal scenario, feature-based methods are high-speed but do not achieve an acceptably low TRE. XFeat establishes the optimal Pareto-efficient point across all three metrics, delivering a 4.01× reduction in error relative to Faber while maintaining near-instantaneous execution and minimal energy demand. Although MINIMA LoFTR-GPU achieves an absolute accuracy ceiling of 5.83px, its prohibitive execution time of 29.01s and energy consumption of 143.22J render it unsuitable for on-board tasks.

Insight #1 Our characterization confirms that no architecture dominates the entire Pareto frontier. We observe a fundamental specialization gap: feature-based methods, though efficient, exhibit low accuracy in multimodal scenarios. Deep learning models bridge this gap, but at the cost of energy inefficiency. Notably, intensity-based methods demonstrate unique resilience, providing stable alignment where others fail, thus confirming a "specialization trade-off" in IR.

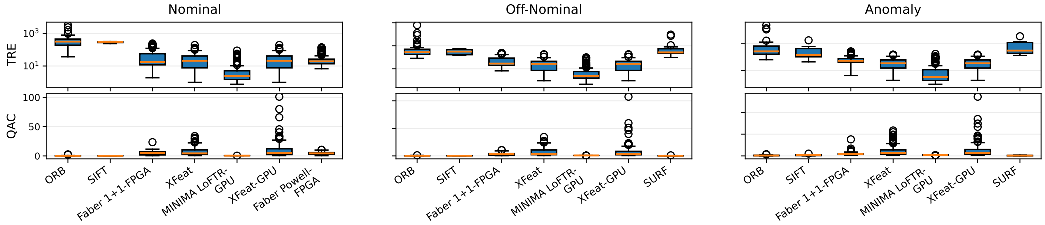


Fig. 8. The boxplots represent the distributions of TRE (top row) and QAC (bottom row) for the Pareto-optimal configurations identified in each scenario across the three operational tiers: **Nominal** (I), **Off-Nominal** (II), and **Anomaly** (III). The data highlight the robust, deterministic performance of intensity-based methods (Faber), which maintain compact distributions across all stress levels, in contrast to the high variance and functional collapse of feature-based descriptors (ORB, SIFT). The QAC metric effectively captures the trade-off between precision and resource efficiency. It assigns the energy-intensive MINIMA methods the lowest QAC scores (the lowest orange bar) across all scenarios, despite their superior sub-pixel accuracy.

4.2 *STAR-Bench*- Parallax Stability

This section analyzes the evolution of the Pareto frontier across the three geometric deformation tiers—Nominal (I), Off-Nominal (II), and Anomaly (III)—specifically within the Multimodal (3) sensing domain[6]. Figure 8 illustrates the distribution of TRE and the resulting QAC across these stress levels. Under **Nominal** orbital conditions (I), a pronounced divergence between registration accuracy and system efficiency is observed. While MINIMA LoFTR (GPU) achieves the lowest median TRE, its QAC remains low, penalized by the high energy cost of dense transformer-based matching on embedded hardware. Conversely, XFeat and Faber (utilizing the Powell optimizer in their FPGA configurations) stand as the best solutions. The former attains a higher but highly variable QAC (5.73 ± 4.85), while the latter features a slightly lower, but more stable, QAC (4.41 ± 0.67). Given XFeat stability issues, the intensity-based method as Faber stands as better for the nominal scenario. In contrast, feature-based methods such as ORB and SIFT exhibit complete functional collapse even in nominal multimodal settings, failing the registration (TRE distributions exceeding the image size). As the deformation level increases toward the **Anomaly** scenario (III), the performance landscape undergoes a significant shift in distribution. For most pipelines, the TRE variance increases dramatically, indicating a loss of deterministic behavior under severe distortions. Notably, XFeat emerges in this tier as a critical "resilience baseline". Although it exhibits a significant number of high-error outliers, its median QAC remains competitive, suggesting that CNN-based models offer a more predictable failure mode than complex transformers under extreme distortion. Crucially, Faber with Powell remains stable across all deformation levels, maintaining minimal TRE variance. This highlights the inherent robustness of global intensity-based optimization strategies: when local keypoint matching fails, hardware-accelerated intensity methods recover high, stable QAC values, which are the only consistent solutions across all deformation tiers.

Insight #2 No single method consistently dominates across all deformation tiers; intensity-based methods like Faber offer remarkable stability and high QAC, but needs to be replaced by deep-learning method, such as XFeat, as the system drifts toward Anomaly (III) regimes. This performance inversion underscores the critical necessity of adaptive selection logic to maintain operational reliability across the entire deformation spectrum.

---

[6] Comprehensive results for the remaining sensing domains are deferred to Appendix B.

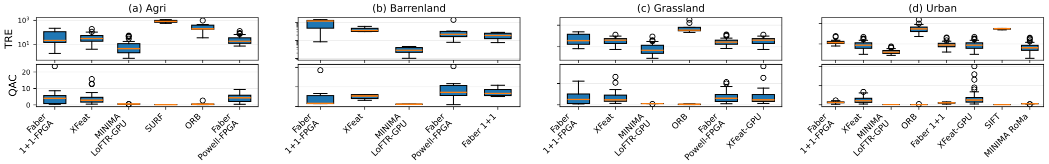


Fig. 9. TRE (top) and QAC (bottom) distributions for the Pareto-optimal configurations in four distinct landscapes: (a) Agricultural ⓐ, (b) Barrenland ⓑ, (c) Grassland ⓒ, and (d) Urban ⓓ. To isolate environmental effects from geometric distortions, all tests are conducted under Nominal (Ⓘ) conditions within the Multimodal (③) sensing domain. Structured urban landscape favours feature matchers such as XFeat. In the other cases, Faber stability leads to higher QAC, further highlighting the specialization trade-off.

4.3 *STAR-Bench*- Analysis Across Terrain Variability

This section evaluates how variations in environmental context, ranging from the high textural complexity of Urban (ⓓ) areas to the relative homogeneity of Grasslands (ⓒ), shift the Pareto frontier, significantly modulating the latency-accuracy-energy trade-off for Multimodal (③) registration under Nominal (Ⓘ) deformations[7]. As illustrated in Figure 9 in Urban (ⓓ) terrains, the presence of well-defined architectural structures and sharp geometric boundaries provides a favorable landscape for deep-learning-based methods such as XFeat and MINIMA LoFTR. Specifically, XFeat achieves its highest operational efficiency in these zones, reaching an QAC of 12.95 with a mean TRE of 10.25px. Due to the high computational latency of the most accurate method, MINIMA LoFTR-GPU ($QAC = 0.52$), XFeat outperforms it in system-level efficiency by a factor of 24.9×. Conversely, in Agriculture (ⓐ), Barrenland (ⓑ), and Grassland (ⓒ), the reduction in high-frequency structural features and the prevalence of repetitive textures significantly degrade standalone descriptor performance. In these scenarios, the Faber family (particularly the hardware-accelerated Powell-FPGA and software 1+1 versions) represents the state of the art for QAC. In Agriculture, Faber Powell-FPGA reaches a QAC of 4.30, outperforming XFeat ($QAC = 4.06$) by approximately 1.1× and surpassing MINIMA LoFTR-GPU ($QAC = 0.55$) by a factor of 7.8×. The challenge is most deep in Barrenland, where Faber 1+1 refinement yields a QAC of 3.66, providing a 1.47× efficiency improvement over XFeat ($QAC = 2.49$). This evidence supports the conclusion that ground content is a primary driver of registration instability, necessitating a terrain-aware framework that dynamically scales between lightweight learned features for urban areas and robust, albeit slower, intensity-based optimizers for homogeneous terrains.

Insight #3 Terrain variability introduces a significant "noise floor" in multimodal registration. This sensitivity confirms that on-board systems cannot rely on a static "best" method, but must dynamically adapt to ground content to ensure consistent mission reliability.

5 *ETNA*: Unified Mission-Aware Framework Multiple IR Pipelines Integration

Building upon the performance insights and the fundamental "specialization trade-off" identified in Insight #1 (§4.1), and addressing the critical need for adaptive pipelines capable of responding to varying system conditions (Insight #2, §4.2) and terrain diversity (Insight #3, §4.3), we propose *ETNA*. This unified framework integrates disparate IR methods, ranging from deep learning to custom hardware-accelerated optimization, to assess how heterogeneous computation can yield tangible system-level benefits.

---

[7] Additional analyses across all experimental setups are available in Appendix B.

By addressing the functional gap where no single solution can consistently match every input pair, *ETNA* orchestrates the methods described in §3.3. Specifically, it leverages XFeat for monomodal tasks and a specialized, co-designed intensity-based module for the multimodal domain to mitigate the inherent weaknesses identified in our earlier characterization (§4.1). This is achieved through the hardware-software optimization explored in §5.2, ensuring that even the most challenging registration tasks remain within the mission's temporal and energetic envelopes. *ETNA* further investigates the feasibility of hybrid registration pipelines. This strategy has demonstrated significant accuracy gains in other domains such as medical imaging [2, 58, 60], supporting its viability within the stringent latency-accuracy-energy trade-off of EO missions.

5.1 *ETNA* Dynamic Pipeline

Rather than relying on monolithic software solutions, *ETNA* introduces a suite of orchestrated paths that leverage a custom-engineered, hardware-native optimization, representing a significant architectural departure from standard intensity-based methods by offloading the most compute-intensive kernels to reconfigurable logic. These pipelines are engineered through four distinct architectural paths, illustrated in Figure 4:

ETNA-Intensity-to-DL (❶): this experimental hybrid pipeline investigates the synergy between the structural robustness of the intensity-based method and the precision of the deep-learning method. It uses our hardware-accelerated engine to perform an initial coarse-grained global alignment, bringing the images into an initial coarse alignment frame at near-instant speeds. The intensity-based stage provides a stable initialization for a deep-learning module (e.g., XFeat) to perform fine-grained refinement, thereby preventing the functional collapse typically observed under large initial geometric offsets. In our results, these hybrid configurations follow the naming convention *ETNA*-[Intensity optimizer]-[DL model]. For instance, *ETNA*-Powell-XFeat denotes a pipeline in which the Powell optimizer serves as the intensity-based initialization, which is subsequently refined by the XFeat deep-learning module.

ETNA-Inverted (❷): this "refinement-first" configuration inverts the hierarchy to address the sub-pixel accuracy limitations of deep-learning methods. It employs deep learning modules for initial coarse alignment and subsequently uses our hardware-accelerated engine for final refinement. This is designed for scenarios where the sensor degrades the precision of keypoint matching. In our results, these hybrid configurations follow the naming convention *ETNA*-Inv-[DL model]-[Intensity optimizer]. For instance, *ETNA*-Inv-XFeat-Powell denotes a pipeline in which XFeat serves initialization, which is subsequently refined by Powell optimizer.

ETNA-Multi (❸): specifically designed for high-stress radiometric domain gaps (③, ④), this path relies exclusively on our hardware-accelerated intensity optimization, bypassing feature matching to maximize reliability.

ETNA-Mono (❹): optimized for efficiency in intra-band scenarios (①, ②), this path leverages XFeat for near-instantaneous processing with minimal energy overhead.

ETNA employs a modality selector to dispatch image pairs to the optimal path based on sensor type, terrain, and expected geometric stress. Using the *STAR-Bench* characterization results, the selector maps acquisition scenarios to the most efficient pipeline. Additionally, we characterize both hybrid pipeline configurations to evaluate their behavior within the high-constraint latency-accuracy-energy trade-off space of on-board satellite systems. This investigation seeks to determine whether the performance gains of multi-stage strategies observed in other domains can be effectively translated to the EO context, potentially bridging functional gaps where monolithic methods fail.

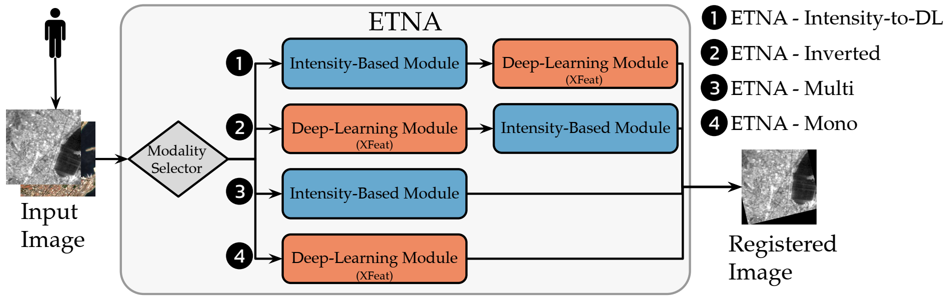


Fig. 10. Architectural overview of the *ETNA* framework. The system receives an input image pair and utilizes a Modality Selector to orchestrate execution across four specialized paths. Path ❶ (*ETNA*-Intensity-to-DL) and Path ❷ (*ETNA*-Inverted) represent hybrid configurations in which intensity-based optimization and deep-learning feature matching are combined in different precedence orders to improve resilience. Path ❸ (*ETNA*-Multi) provides a direct route for challenging multimodal scenarios, while Path ❹ (*ETNA*-Mono) leverages lightweight CNN-based methods for efficient monomodal registration. This modular design enables the framework to adapt dynamically to varying sensing modalities and geometric conditions.

5.2 *ETNA*-Multi: an Hardware/Software Co-Design for Multimodal Registration

While intra-band (①), monomodal (②), and multi-polarity (④) scenarios appear suitable for on-board execution, multimodal registration remains the most challenging scenario (Insight #2, §4.2). To push the operational boundaries of multimodal (③) registration, we introduce ***ETNA*-Multi**. This hardware/software co-design architecture targets the AMD Kria, optimizing the intensity-based pipeline for the limited computational budget of COTS hardware.

Pyramidal Optimization and Registration - Inspired by hierarchical multiresolution techniques [11, 39, 52], we integrated the intensity-based registration into a pyramidal architecture (Figure 11). The input image pair is decomposed into N resolution levels, ranging from 32×32 to 256×256 pixels. To maximize efficiency, we redesigned the control logic of the Powell and 1+1 optimizers to employ resolution-dependent search parameters. At the coarsest level (32×32), a wide search range is used to capture large translational drifts within the smoothed image landscape. As the process moves toward finer resolutions, the search space is progressively constrained around the current estimate. The optimal transformation parameters T_a^* identified at level L_i are scaled up to serve as the initialization for level L_{i+1}, thereby eliminating redundant exploration at higher scales. To further optimize the energy-latency product, we implement a level-specific early-exit strategy. The system advances to the next pyramid level when it reaches a stagnation threshold, defined as a fixed number of iterations without significant improvement in the similarity metric.

Runtime-Configurable Hardware Acceleration - A critical limitation of standard FPGA accelerators is their rigidity: they are typically synthesized for fixed configurations (e.g., fixed image resolutions), requiring a complete bitstream reconfiguration that can take seconds and nullifying acceleration benefits in latency-sensitive scenarios such as the one we target in this work. The accelerators available within Faber [13], which offload the CC and MI similarity metrics to the FPGA, feature such a rigid approach, making them unsuitable for our purposes. For this reason, we redesigned the original High-Level Synthesis (HLS) code [14] of these accelerators to support any image resolution up to a fixed maximum size defined at design time, shifting the rigidity to this parameter and providing users with greater freedom. In particular, this new version of the accelerators now takes as input not only the images, whose similarity has to be computed, but also their resolutions, and propagates this information through the internal stages of the accelerator

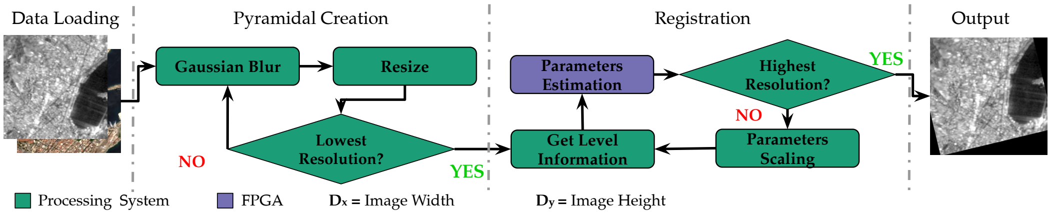


Fig. 11. Hierarchical multi-resolution registration workflow. The process utilizes a Gaussian pyramidal decomposition to perform iterative coarse-to-fine alignment. By initiating parameter estimation at the user-selected lower resolution and scaling the identified transformation to the target resolution, the system effectively reduces the search space for the heuristic optimizer, thereby minimizing the total number of high-resolution iterations and hardware invocations.

pipelines, which now operate and iterate based on these external parameters supplied by the user at runtime. Finally, we integrated these accelerators within *ETNA*-Multi.

This design upgrade ensures the hardware context processes images up to 512 × 512 without reloading bitstreams, with peak critical resource (i.e., BRAM) usage at 47.6% (see Appendix B.2 for more details). By exposing the image resolution, the software orchestration layer can dynamically resize the active processing window at each pyramid level without reloading the FPGA bitstream. Consequently, *ETNA*-Multi executes the entire coarse-to-fine pyramidal registration stack as a continuous stream, reducing the inter-level transition time and enabling the entire process to be deployed for real-time EO missions.

6 *ETNA* Evaluation and Analysis

This section evaluates the performance of the *ETNA* framework on the *STAR-BENCH* characterization, quantifying its impact on registration efficiency and accuracy. We first analyze the speedup and energy savings achieved by the *ETNA*-Multi pyramidal architecture, and then discuss the overall system-level improvements across all sensing modalities.

6.1 *ETNA*- Analysis Across Data Types

This subsection integrates the proposed *ETNA* method into the Pareto analysis introduced in §4.1. Figure 12 and Table 2 compares its performance across the same sensing domains and deformation tiers, showing that *ETNA* effectively redefines the Pareto frontier, improving the current state-of-the-art in terms of QAC.

Intra-band Registration (①). In the sensing domain where §4.1 identified a critical void between unstable feature-based methods and prohibitive deep-learning models, *ETNA* introduces the ETNA-Powell-XFeat (Path ❶). This hybrid pipeline achieves a mean TRE of 1.98px in only 0.91s. While the hybrid approach maintains an accuracy profile comparable to standalone XFeat (1.59px), it delivers a transformative 18× performance improvement over standalone intensity-based methods, such as Faber 1+1-FPGA (36.19px), effectively stabilizing the registration process without the latency penalties of heavier models. Although MINIMA LG remains the most accurate baseline (0.22px), its 60.86s latency violates the real-time threshold, positioning ETNA-Powell-XFeat and XFeat as the only viable candidates for sub-second missions requiring high precision.

Monomodal Registration (②). In this scenario, the Pareto-optimal frontier found in the characterization (§4.1) effectively surpasses the performance of all methods. Notably, seven of the twenty evaluated *ETNA* pipelines successfully reside within the operationally viable region, defined by the mission-critical latency and accuracy thresholds illustrated in §3.3.

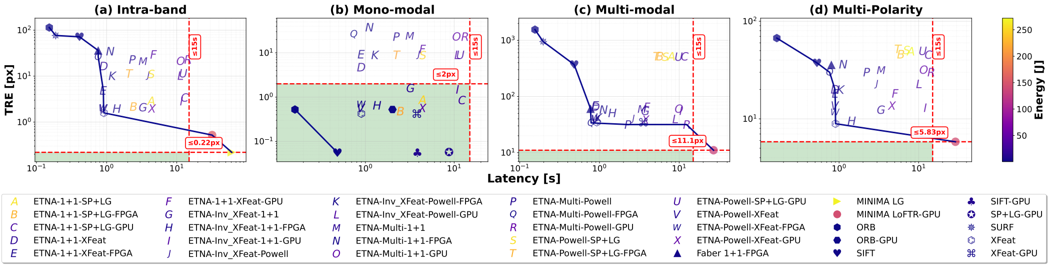


Fig. 12. The latency-accuracy-energy trade-off Pareto frontier for (a) Intra-band, (b) Monomodal, (c) Multimodal, and (d) Multi-polarity registration. Markers A-X denote *ETNA* pipeline configurations executed on both CPU, FPGA, and GPU architectures. This Pareto line shows that *ETNA* configurations consistently occupy the latency-accuracy-energy frontier in the challenging multimodal domains (c). The numerical values corresponding to the Pareto frontier are reported in Table 2. To ensure clarity, this figure displays only the Pareto-optimal baselines and the proposed *ETNA* configurations, omitting non-optimal methods.

Table 2. Comparison of Pareto Front Pipelines across sensing modalities. ① Intra-band, ② Monomodal, ③ Multimodal, ④ Multi-Polarity. **Bold** entries highlight the best performance in each group.

Type	Pipeline	Lat. [s]	TRE [px]	En. [J]	QAC
①	ORB	**0.15 ± 0.07**	114.1 ± 1384.6	**0.37 ± 0.24**	0.87
	SURF	0.19 ± 0.09	76.7 ± 1160.3	0.54 ± 0.34	1.30
	SIFT	0.41 ± 0.18	71.7 ± 909.3	1.44 ± 0.68	1.37
	Faber 1+1-FPGA	0.76 ± 0.05	36.2 ± 30.9	2.48 ± 0.14	2.59
	ETNA-Multi-Powell-FPGA	0.76 ± 0.49	26.2 ± 18.7	2.15 ± 0.22	3.53
	ETNA-1+1-XFeat-FPGA	0.90 ± 0.06	4.88 ± 39.8	6.19 ± 0.70	9.03
	ETNA-Powell-XFeat-FPGA	0.91 ± 0.07	2.05 ± 12.9	4.41 ± 0.34	15.5
	ETNA-Powell-XFeat	0.91 ± 0.07	1.98 ± 11.2	10.9 ± 1.34	7.75
	XFeat	0.92 ± 0.11	1.59 ± 10.00	2.15 ± 0.41	**26.7**
	MINIMA LoFTR-GPU	31.9 ± 9.30	0.52 ± 1.82	157.2 ± 42.5	0.63
	MINIMA LG	60.9 ± 2.99	**0.22 ± 0.49**	267.7 ± 26.7	0.37
②	ORB	**0.16 ± 0.07**	0.53 ± 5.72	**0.38 ± 0.25**	**109.9**
	SIFT	0.48 ± 0.21	**0.06 ± 0.13**	1.70 ± 0.82	56.8
③	ORB	**0.16 ± 0.06**	1548.4 ± 6383.7	**0.50 ± 0.21**	0.06
	SURF	0.20 ± 0.09	947.3 ± 1797.8	0.72 ± 0.35	0.11
	SIFT	0.48 ± 0.19	369.8 ± 433.7	1.92 ± 0.77	0.27
	Faber 1+1-FPGA	0.77 ± 0.18	60.6 ± 59.4	2.56 ± 0.87	1.58
	ETNA-Multi-Powell-FPGA	0.79 ± 0.57	33.9 ± 28.0	2.13 ± 0.22	**2.78**
	ETNA-Multi-Powell	2.28 ± 0.26	31.7 ± 25.5	9.38 ± 1.44	2.44
	ETNA-Multi-Powell-GPU	12.4 ± 10.4	31.7 ± 25.5	46.9 ± 42.1	1.27
	MINIMA LoFTR-GPU	27.0 ± 7.35	**11.1 ± 22.9**	133.3 ± 32.7	0.69
④	ORB	**0.16 ± 0.06**	67.2 ± 331.8	**0.47 ± 0.17**	1.48
	SIFT	0.53 ± 0.22	37.4 ± 385.1	1.93 ± 0.74	2.54
	ETNA-Multi-Powell-FPGA	0.76 ± 0.49	30.0 ± 23.5	2.13 ± 0.23	3.12
	ETNA-1+1-XFeat	0.89 ± 0.04	20.1 ± 54.4	15.2 ± 3.04	2.83
	ETNA-Powell-XFeat	0.90 ± 0.07	13.7 ± 42.6	10.7 ± 1.25	4.11
	ETNA-Powell-XFeat-FPGA	0.90 ± 0.07	11.8 ± 33.7	4.36 ± 0.34	6.19
	XFeat	0.90 ± 0.09	8.92 ± 24.8	2.12 ± 0.28	**9.06**
	MINIMA LoFTR-GPU	29.0 ± 8.53	**5.83 ± 16.9**	143.2 ± 38.2	0.67

Multimodal Registration (③). The most significant impact of the *ETNA* co-design is observed in the multimodal domain, which was previously a victim of the "specialization trade-off" for standalone algorithms exhibiting mean errors exceeding 360px. In this scenario, *ETNA*-Multi-Powell-FPGA (Marker Q, Path ❸) redefines the efficiency frontier by achieving a mean TRE of 33.9px in just 0.79s. This result represents a 1.79x improvement in accuracy relative to non-optimized intensity-based versions. While deep-learning models such as MINIMA achieve lower absolute error (11.1px), their prohibitive latency (> 26s) and high energy cost (133.3J) violate the real-time processing deadlines of EO missions. Consequently, the 33.9px TRE achieved by *ETNA*-Multi constitutes the current Pareto-optimal frontier for autonomous on-board processing. These results validate

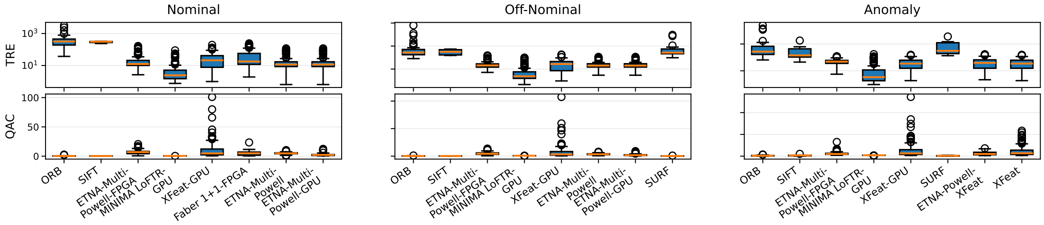


Fig. 13. System-level stability and resilience analysis for Multimodal (③) registration. The boxplots illustrate the distribution of TRE [px] (top row) and QAC (bottom row) for the Pareto-optimal configurations identified in each scenario across the three experimental deformation tiers Ⅰ-Ⅲ. Results are aggregated across the four terrains. The distributions highlight the functional collapse of feature-based methods, which exhibit high errors even in Nominal conditions. Conversely, *ETNA* pipelines and deep-learning method benchmarks maintain operational survival and achieve significantly higher QAC scores.

the efficacy of hardware-accelerated intensity optimization in bridging the radiometric domain gap under strict real-time constraints. Notably, no hybrid registration pipelines lie on the Pareto optimality frontier for this modality, indicating that the extreme radiometric differences between Optical and SAR data favor specialized single-method approaches over sequential combinations.

Multipolarization Registration (④). The results are closely aligned with the performance trends observed in the Intra-band (①) scenario. In this modality, *ETNA* successfully populates the Pareto frontier, introducing the ETNA-Powell-XFeat-FPGA configuration (Marker T, Path ❶) as a robust alternative to the standalone XFeat baseline. While XFeat remains a strong candidate for high-speed registration, the ETNA-Powell-XFeat-FPGA configuration offers a more stable trade-off, exhibiting significantly less variation in the QAC metric across test samples.

6.2 *ETNA*- Parallax Stability

After benchmarking existing architectures in §4.2, we now evaluate how *ETNA* redefines the latency-accuracy-energy trade-off in the same experimental setup.

As illustrated in Figure 13, under Nominal (Ⅰ) conditions, the framework establishes a high stability baseline where classical methods like ORB fail to provide meaningful convergence, exhibiting a mean TRE of 500.43px. In contrast, the ETNA-Multi-Powell-FPGA (❸) configuration sets the group's stability benchmark with a QAC of 6.58 and a mean TRE of 22.69px in only 0.76s. While the deep-learning baseline MINIMA LoFTR-GPU achieves higher absolute precision (7.28px), its QAC of 0.54 entails prohibitive computational overhead, rendering the *ETNA* configuration approximately 11× more efficient in terms of QAC. The architectural resilience of *ETNA* becomes even more critical as the system transitions to Off-Nominal (Ⅱ) and Anomaly (Ⅲ) stress levels. In the Anomaly tier, ETNA-Multi-Powell-FPGA maintains a TRE of 50.51px with a stable QAC of 2.34. In comparison, the XFeat baseline exhibits the highest anomaly resilience, with an QAC of 4.86 and a TRE of 41.98px. Notably, across all stress levels, *ETNA* configurations consistently reside within the sub-second latency region. In contrast, MINIMA LoFTR-GPU experiences significant latency fluctuations (from 37.03s in Nominal to 21.37s in Anomaly) while never surpassing a QAC of 0.83.

Therefore, dynamically selecting between hardware-accelerated intensity optimizers and lightweight deep-learning methods, *ETNA* prevents both the efficiency collapse typical of heavy deep-learning-based methods, ensuring consistent mission reliability across the entire mission life cycle, with the highest accuracy for both Nominal and Off-Nominal deformations.

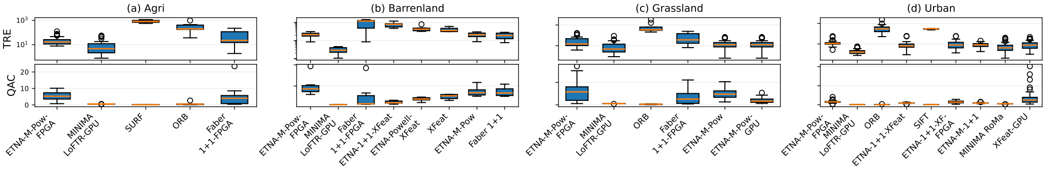


Fig. 14. The plots illustrate the registration TRE (top row) and QAC (bottom row) for the Pareto-optimal configurations identified in each scenario across the four representative terrains: (a) Agri (ⓐ), (b) Barrenland (ⓑ), (c) Grassland (ⓒ), and (d) Urban (ⓓ). Notably, hybrid configurations such as ETNA-Powell-XFeat (❶) and ETNA-1+1-XFeat (❶) demonstrate superior resilience in low-texture environments, such as Barrenland, maintaining high QAC and low TRE variance, whereas monolithic feature-based methods exhibit functional instability. Furthermore, the hardware-accelerated intensity-based paths (❸) achieve the highest group efficiency across multimodal scenarios, validating the benefit of offloading critical similarity kernels to reconfigurable logic to optimize the latency-accuracy-energy trade-off.

6.3 *ETNA*- Analysis Across Terrain Variability

Building directly on the environmental sensitivities identified in §4.3, this final evaluation dimension investigates how terrain heterogeneity impacts the operational stability of the *ETNA* framework.

Figure 14 describes the analysis of different terrains through *STAR-Bench*. In Urban terrain (ⓓ), the presence of distinct man-made structures provides a favorable landscape for feature-based approaches. The ETNA-1+1-XFeat-FPGA configuration (Path ❶) achieves performance approximately parity with standalone XFeat, reaching a mean TRE of 11.58px (due to FPGA approximation, the CPU version achieves the same TRE) while significantly reducing QAC variance. Notably, while MINIMA LoFTR-GPU achieves an absolute accuracy of 1.85px, its QAC remains stagnant at 0.52. The prohibitive 28.28s latency of the model renders it 14.6× less efficient than the *ETNA* alternatives, which maintain sub-second response times. In more challenging environments, such as Agriculture (ⓐ) and Grassland (ⓒ), the prevalence of repetitive textures and lower textural complexity increases the mean error across all pipelines. In Agriculture, the ETNA-Multi-Powell-FPGA (Path ❸) achieves the best QAC of 5.46 with a TRE of 26.91px. Similarly, in Grassland, this configuration provides stability with an QAC of 6.09, thereby maintaining reliability when classical descriptors typically fail. The system encounters its most severe environmental bottleneck in Barrenland (ⓑ), characterized by extreme radiometric homogeneity. In this biome, ETNA-Multi-Powell-FPGA remains the most balanced option, with a QAC of 3.71 ± 1.91 and a TRE of 32.56px. Although MINIMA LG can achieve a precision of 5.30px, its 34.91s execution time yields the lowest efficiency score ($QAC = 0.59$) for this terrain. Across all biomes, the data confirms that *ETNA* successfully populates the Pareto frontier, providing a suite of terrain-aware configurations capable of surviving the radiometric domain gap.

Therefore, by dynamically selecting between Deep-Learning-based methods (Path ❹) for rich-texture areas and intensity optimizers (Path ❸) for repetitive terrain, the framework employs hybrid pipelines to achieve an optimal trade-off across all evaluated biomes.

7 Related Work

The acceleration of IR for resource-constrained platforms has followed two divergent paths: (i) hardware-centric optimizations that overlook end-to-end precision, and (ii) algorithmic innovations evaluated on high-power hardware. Our work bridges this divide with *STAR-Bench* for standardized evaluation and *ETNA* for a balanced, high-performance architecture.

Hardware and Algorithmic Constraints. Hardware-focused studies, such as Du et al. [21], validate FPGA offloading but rely on proxy metrics, such as matching ratio rather than end-to-end TRE, leaving downstream registration fidelity unverified. Conversely, algorithmic proposals often lack systems-level rigor; for instance, Inzerillo et al. [31] omit power metrics, while Meoni et al. [38] evaluate DNNs on 300W GPUs, significantly exceeding CubeSat power budgets. Recently, Priyasada et al. [43] introduced a framework for onboard registration that employs two-stage knowledge distillation on the Intel Myriad VPU. While achieving impressive model sizes, their approach focuses on intra-band (①) alignment, whereas *ETNA* addresses the more complex problem of multimodal registration under real-world geometric deformations and sensing modalities.

On-Board Applications and Robustness. Application-level works in change detection, as Earth+ [20] and Kodan [15], assume perfectly aligned input imagery, overlooking the computational bottleneck of the prerequisite registration step. While REOBench [34] has established a benchmark for foundation-model robustness against environmental and geometric perturbations, it focuses on application-level tasks, such as segmentation. *STAR-Bench* extends this diagnostic rigor to the system level, characterizing the prerequisite registration phase under realistic on-board constraints. This review proves a critical gap in the literature, systematically addressed by our work.

8 Conclusion and Future Work

This paper addresses a fundamental prerequisite for autonomous Earth Observation (EO): enabling reliable on-board Image Registration (IR) under strict system-level constraints. The implications of this research are clear: it aims to eliminate ground dependence to ensure Orbital Edge Computing (OEC). If a satellite pipeline continues to rely on ground-based registration, it is sacrificing autonomy and latency in retrieving insights. To address this, we introduced *STAR-Bench*, a reproducible framework that uses the Quality-Adjusted Cost (QAC) metric to reveal the *"specialization trade-off"*, demonstrating that no single method optimizes the latency-accuracy-energy trade-off across diverse scenarios. Guided by these insights, we proposed *ETNA*, a mission-aware framework that orchestrates heterogeneous IR pipelines. By populating the Pareto frontier, *ETNA* achieved up to 12.03× QAC gain over state-of-the-art baselines while maintaining sub-second performance on Commercial-Off-The-Shelf (COTS) hardware.

While *ETNA* and *STAR-Bench* establish a robust foundation, certain **limitations** persist. Regarding data representativeness, although our Monte Carlo augmentation rigorously models different scenarios, unforeseen ones may still occur. Secondly, concurrent OEC tasks may increase a resource contention requiring further analysis. As **future work**, we will address the presented limitations and extend *STAR-Bench* to hyperspectral workload, possibly automating *ETNA*'s pipeline dispatching.

In conclusion, our work demonstrates that COTS hardware is no longer the primary bottleneck for OEC; rather, the true limitation lies in the lack of adaptive frameworks. By bridging this gap, *ETNA* and *STAR-Bench* lay the foundation for deploying autonomous on-board vision systems, reducing ground dependency and enabling satellite autonomy.

Acknowledgments

The authors used an LLM to improve text readability; all output was manually verified for technical correctness. This work was supported by the European Commission in the context of the HORIZON EIC Cancerscan project (GA 101186829). This work was supported in part by the AMD University Program. The authors are grateful for the anonymous reviewers' and shepherd feedback.

References

[1] 2025. Copernicus Browser. https://browser.dataspace.copernicus.eu/. Accessed: 2025-09-17.

[2] Mohamed Abdel-Basset, Ahmed E Fakhry, Ibrahim El-Henawy, Tie Qiu, and Arun Kumar Sangaiah. 2017. Feature and intensity based medical image registration using particle swarm optimization. *Journal of medical systems* 41, 12 (2017), 197.

[3] European Space Agency. 2020. Low Earth orbit. https://www.esa.int/ESA_Multimedia/Images/2020/03/Low_Earth_orbit. Image, ESA Multimedia.

[4] Moab Arar, Yiftach Ginger, Dov Danon, Amit H Bermano, and Daniel Cohen-Or. 2020. Unsupervised multi-modal image registration via geometry preserving image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 13410–13419.

[5] Guillaume Astruc, Nicolas Gonthier, Clement Mallet, and Loic Landrieu. 2025. AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 19530–19540.

[6] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. Surf: Speeded up robust features. In *European conference on computer vision*. Springer, 404–417.

[7] Nikolaos Ioannis Bountos, Arthur Ouaknine, and David Rolnick. 2023. FoMo-Bench: a multi-modal, multi-scale and multi-task Forest Monitoring Benchmark for remote sensing foundation models. *arXiv preprint arXiv:2312.10114* (2023).

[8] Konstantinos-Panagiotis Bouzoukis, Georgios Moraitis, Vassilis Kostopoulos, and Vaios Lappas. 2025. An Overview of CubeSat Missions and Applications. *Aerospace* 12, 6 (2025), 550.

[9] Lisa Gottesfeld Brown. 1992. A survey of image registration techniques. *ACM computing surveys (CSUR)* 24, 4 (1992), 325–376.

[10] Zhuo Cheng and Brandon Lucia. 2025. Nanosatellite Constellation and Ground Station Co-design for Low-Latency Critical Event Detection. *arXiv preprint arXiv:2503.01756* (2025).

[11] Arlene A Cole-Rhodes, Kisha L Johnson, Jacqueline LeMoigne, and Ilya Zavorin. 2003. Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. *IEEE transactions on image processing* 12, 12 (2003), 1495–1511.

[12] Angela Cratere, Leandro Gagliardi, Gabriel A Sanca, Federico Golmar, and Francesco Dell'Olio. 2024. On-board computer for CubeSats: state-of-the-art and future trends. *IEEE Access* (2024).

[13] Eleonora D'Arnese, Davide Conficconi, Emanuele Del Sozzo, Luigi Fusco, Donatella Sciuto, and Marco D Santambrogio. 2022. Faber: a Hardware/Soft-ware Toolchain for Image Registration. *IEEE Transactions on Parallel and Distributed Systems* (2022), 1–12. doi:10.1109/TPDS.2022.3218898

[14] Emanuele Del Sozzo, Davide Conficconi, Alberto Zeni, Mirko Salaris, Donatella Sciuto, and Marco D Santambrogio. 2022. Pushing the level of abstraction of digital system design: A survey on how to program fpgas. *Comput. Surveys* 55, 5 (2022), 1–48.

[15] Bradley Denby, Krishna Chintalapudi, Ranveer Chandra, Brandon Lucia, and Shadi Noghabi. 2023. Kodan: Addressing the computational bottleneck in space. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*. 392–403.

[16] Bradley Denby and Brandon Lucia. 2019. Orbital edge computing: Machine inference in space. *IEEE Computer Architecture Letters* 18, 1 (2019), 59–62.

[17] Bradley Denby and Brandon Lucia. 2020. Orbital edge computing: Nanosatellite constellations as a new class of computer system. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*. 939–954.

[18] Josip Djolonga, Jessica Yung, Michael Tschannen, Rob Romijnders, Lucas Beyer, Alexander Kolesnikov, Joan Puigcerver, Matthias Minderer, Alexander D'Amour, Dan Moldovan, et al. 2021. On robustness and transferability of convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16458–16468.

[19] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment* 120 (2012), 25–36.

[20] Kuntai Du, Yihua Cheng, Peder Olsen, Shadi Noghabi, and Junchen Jiang. 2025. Earth+: On-board satellite imagery compression leveraging historical earth observations. In *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1*. 361–376.

[21] Xin Du, Cankun Yang, Ruofei Zhong, Qingyang Li, Yuanhang Wang, Zhaoming Huang, and Xianlin Liu. 2024. Hardware-Optimized Architecture of On-Board Registration for Remote-Sensing Images—Take SURF as an Example. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), 8230–8249.

[22] Edge SpAIce project consortium. 2025. Edge SpAIce: Novel Edge-AI system for accurate and near real-time plastic detection and monitoring in marine environment. https://edgespaice.eu/. Accessed: 2025-07-23; EU Horizon Europe project, grant agreement no. 101135358.

[23] Johan Edstedt et al. 2024. RoMa: Robust Dense Matching – Official PyTorch Implementation. https://github.com/Parskatt/RoMa. Accessed: 2024-05-20.

[24] Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, and Michael Felsberg. 2024. Roma: Robust dense feature matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 19790–19800.

[25] Martin A. Fischler and Robert C. Bolles. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24, 6 (June 1981), 381–395. doi:10.1145/358669.358692

[26] Gianluca Furano, Gabriele Meoni, Aubrey Dunne, David Moloney, Veronique Ferlet-Cavrois, Antonis Tavoularis, Jonathan Byrne, Léonie Buckley, Mihalis Psarakis, Kay-Obbe Voss, et al. 2020. Towards the use of artificial intelligence on the edge in space systems: Challenges and opportunities. *IEEE Aerospace and Electronic Systems Magazine* 35, 12 (2020), 44–56.

[27] Gianluca Giuffrida, Lorenzo Diana, Francesco de Gioia, Gionata Benelli, Gabriele Meoni, Massimiliano Donati, and Luca Fanucci. 2020. CloudScout: A Deep Neural Network for On-Board Cloud Detection on Hyperspectral Images. *Remote Sensing* 12, 14 (2020). doi:10.3390/rs12142205

[28] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* (2017). doi:10.1016/j.rse.2017.06.031

[29] Hank Heidt, Jordi Puig-Suari, Augustus Moore, Shinichi Nakasuka, and Robert Twiggs. 2000. CubeSat: A new generation of picosatellite for education and industry low-cost space experimentation. (2000).

[30] Kouta Hirotaki, Shunsuke Moriya, Tsunemichi Akita, Kazutoshi Yokoyama, and Takeji Sakae. 2022. Image preprocessing to improve the accuracy and robustness of mutual-information-based automatic image registration in proton therapy. *Physica Medica* 101 (2022), 95–103. doi:10.1016/j.ejmp.2022.08.005

[31] Gabriele Inzerillo, Diego Valsesia, Aniello Fiengo, and Enrico Magli. 2025. Compress-Align-Detect: onboard change detection from unregistered images. *arXiv preprint arXiv:2507.15578* (2025).

[32] Xiaohua Jia, Tao Lv, Feng He, and Hejiao Huang. 2017. Collaborative data downloading by using inter-satellite links in LEO satellite networks. *IEEE Transactions on Wireless Communications* 16, 3 (2017), 1523–1532.

[33] Jiayuan Li, Pengcheng Shi, Qingwu Hu, and Yongjun Zhang. 2023. RIFT2: Speeding-up RIFT with a new rotation-invariance technique. *arXiv preprint arXiv:2303.00319* (2023).

[34] Xiang Li, Yong Tao, Siyuan Zhang, Siwei Liu, Zhitong Xiong, Chunbo Luo, Lu Liu, Mykola Pechenizkiy, Xiao Xiang Zhu, and Tianjin Huang. 2025. REOBench: Benchmarking Robustness of Earth Observation Foundation Models. *arXiv preprint arXiv:2505.16793* (2025).

[35] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. 2023. LightGlue: Local Feature Matching at Light Speed. In *ICCV*.

[36] David G Lowe. 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 60 (2004), 91–110.

[37] Jiayi Ma, Xingyu Jiang, Aoxiang Fan, Junjun Jiang, and Junchi Yan. 2021. Image matching from handcrafted to deep features: A survey. *International Journal of Computer Vision* 129, 1 (2021), 23–79.

[38] Gabriele Meoni, Roberto Del Prete, Federico Serva, Alix De Beusscher, Olivier Colin, and Nicolas Longépé. 2024. Unlocking the use of raw multispectral earth observation imagery for onboard artificial intelligence. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 17 (2024), 12521–12537.

[39] Behrooz Nasihatkon and Fredrik Kahl. 2017. Multiresolution search of the rigid motion space for intensity-based registration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 1 (2017), 179–191.

[40] Sourabh Paul and Umesh C Pati. 2021. A comprehensive review on remote sensing image registration. *International Journal of Remote Sensing* 42, 14 (2021), 5396–5432.

[41] Planet Labs PBC. 2024. *Planet Fusion Monitoring Technical Specification: Surface Reflectance.* Technical Report. Planet Labs PBC, San Francisco, CA, USA. https://support.planet.com/hc/en-us/article_attachments/19693522324765

[42] Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, and Erickson R Nascimento. 2024. Xfeat: Accelerated features for lightweight image matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2682–2691.

[43] Darshana Priyasad, Tharindu Fernando, Maryam Haghighat, Harshala Gammulle, Roberto Del Prete, and Clinton Fookes. 2026. Two-stage offline knowledge distillation for onboard registration of multispectral satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing* 232 (2026), 319–335.

[44] Jiangwei Ren, Xingyu Jiang, Zizhuo Li, Dingkang Liang, Xin Zhou, and Xiang Bai. 2025. MINIMA: Modality Invariant Image Matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[45] requiemonk. 2024. Sentinel-1/2 Image Pairs Segregated by Terrain. https://www.kaggle.com/datasets/requiemonk/sentinel12-image-pairs-segregated-by-terrain. Accessed: 2026-01-13.

[46] Mitchell Roznik, Milton Boyd, and Lysa Porth. 2022. Improving crop yield estimation by applying higher resolution satellite NDVI imagery and high-resolution cropland masks. *Remote Sensing Applications: Society and Environment* 25

(2022), 100693.

[47] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. 2011. ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision*. Ieee, 2564–2571.

[48] M. Schmitt, L. H. Hughes, and X. X. Zhu. 2018. THE SEN1-2 DATASET FOR DEEP LEARNING IN SAR-OPTICAL DATA FUSION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1 (2018), 141–146. doi:10.5194/isprs-annals-IV-1-141-2018

[49] Daniel Selva and David Krejci. 2012. A survey and assessment of the capabilities of Cubesats for Earth observation. *Acta Astronautica* 74 (2012), 50–68.

[50] Nongmeikapam Thoiba Singh, Charnpreet Kaur, Amrita Chaudhary, and Shefali Goyal. 2023. Preprocessing of Medical Images using Deep Learning: A Comprehensive Review. In *2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*. 521–527. doi:10.1109/ICAISS58487.2023.10250462

[51] Chuanhao Sun, Yu Zhang, Bill Tao, Deepak Vasisht, and Mahesh Marina. 2025. DeepSpace: Super Resolution Powered Efficient and Reliable Satellite Image Data Acquistion. In *Proceedings of the ACM SIGCOMM 2025 Conference*. 311–328.

[52] Philippe Thévenaz and Michael Unser. 2000. Optimization of mutual information for multiresolution image registration. *IEEE transactions on image processing* 9, 12 (2000), 2083–2099.

[53] UP42. 2025. *Polarization*. UP42 Documentation. https://docs.up42.com/data/polarization Accessed: 2025-11-20.

[54] Deepak Vasisht, Jayanth Shenoy, and Ranveer Chandra. 2021. L2D2: Low latency distributed downlink for LEO satellites. In *Proceedings of the 2021 ACM SIGCOMM 2021 Conference*. 151–164.

[55] Paul Viola and William M Wells III. 1997. Alignment by maximization of mutual information. *International journal of computer vision* 24, 2 (1997), 137–154.

[56] Roger Walker, David Binns, Cristina Bramanti, Massimo Casasco, Paolo Concari, Dario Izzo, Davar Feili, Pablo Fernandez, Jesus Gil, Philipp Hager, Detlef Koschny, Vasco Pesquita, Neil Wallace, Ian Carnelli, Michael Khan, Mehdi Scoubeau, and Daniela Taubert. 2018. Deep-space CubeSats: thinking inside the box. *Astronomy & Geophysics* 59, 5 (Oct. 2018), 5–24. doi:10.1093/astrogeo/agy039

[57] R Wilkinson, MM Mleczko, RJW Brewin, KJ Gaston, M Mueller, JD Shutler, X Yan, and K Anderson. 2024. Environmental impacts of earth observation data in the constellation and cloud computing era. *Science of The Total Environment* 909 (2024), 168584.

[58] Marek Wodzinski, Niccolò Marini, Manfredo Atzori, and Henning Müller. 2024. RegWSI: Whole slide image registration using combined deep feature-and intensity-based methods: Winner of the ACROBAT 2023 challenge. *Computer methods and programs in biomedicine* 250 (2024), 108187.

[59] Shufan Wu, Wen Chen, Caixia Cao, Chuanxin Zhang, and Zhongcheng Mu. 2021. A multiple-CubeSat constellation for integrated earth observation and marine/air traffic monitoring. *Advances in Space Research* 67, 11 (2021), 3712–3724.

[60] Jiucheng Xie, Chi-Man Pun, Zhaoqing Pan, Hao Gao, and Baoyun Wang. 2019. Automatic medical image registration based on an integrated method combining feature and area information. *Neural Processing Letters* 49, 1 (2019), 263–284.

[61] Mengqiu Xu, Kaixin Chen, Heng Guo, Yixiang Huang, Ming Wu, Zhenwei Shi, Chuang Zhang, and Jun Guo. 2025. MFogHub: Bridging Multi-Regional and Multi-Satellite Data for Global Marine Fog Detection and Forecasting. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 12637–12646.

[62] Siyue Zhang, Yiyong Xiao, Pei Yang, Yinglai Liu, Wenbing Chang, and Shenghan Zhou. 2019. An effectiveness evaluation model for satellite observation and data-downlink scheduling considering weather uncertainties. *Remote Sensing* 11, 13 (2019), 1621.

[63] Xinyue Zhang, Chengcai Leng, Yameng Hong, Zhao Pei, Irene Cheng, and Anup Basu. 2021. Multimodal remote sensing image registration methods and advancements: A survey. *Remote Sensing* 13, 24 (2021), 5128.

[64] Xiyang Zhi, Junhua Yan, Yiqing Hang, and Shunfei Wang. 2019. Realization of CUDA-based real-time registration and target localization for high-resolution video images. *Journal of Real-Time Image Processing* 16, 4 (2019), 1025–1036.

[65] Bai Zhu, Liang Zhou, Simiao Pu, Jianwei Fan, and Yuanxin Ye. 2023. Advances and challenges in multimodal remote sensing image registration. *IEEE Journal on Miniaturization for Air and Space Systems* 4, 2 (2023), 165–174.

[66] Barbara Zitova and Jan Flusser. 2003. Image registration methods: a survey. *Image and vision computing* 21, 11 (2003), 977–1000.

A APPENDIX

A.1 Geographic Distribution

To ensure the generalizability of the results, the dataset used for the *STAR-Bench* characterization phase comprises image pairs acquired from diverse global regions. The selection of sites was driven by the requirement to cover a wide array of terrestrial biomes, each presenting distinct challenges for image registration. Locations were selected to encompass a range of latitudes, illumination conditions, and seasonal changes. Sites range from high-texture environments, such as Urban (ⓓ) and Agricultural (ⓐ) areas, to extremely critical and homogeneous scenarios, such as Barrenland (ⓑ).

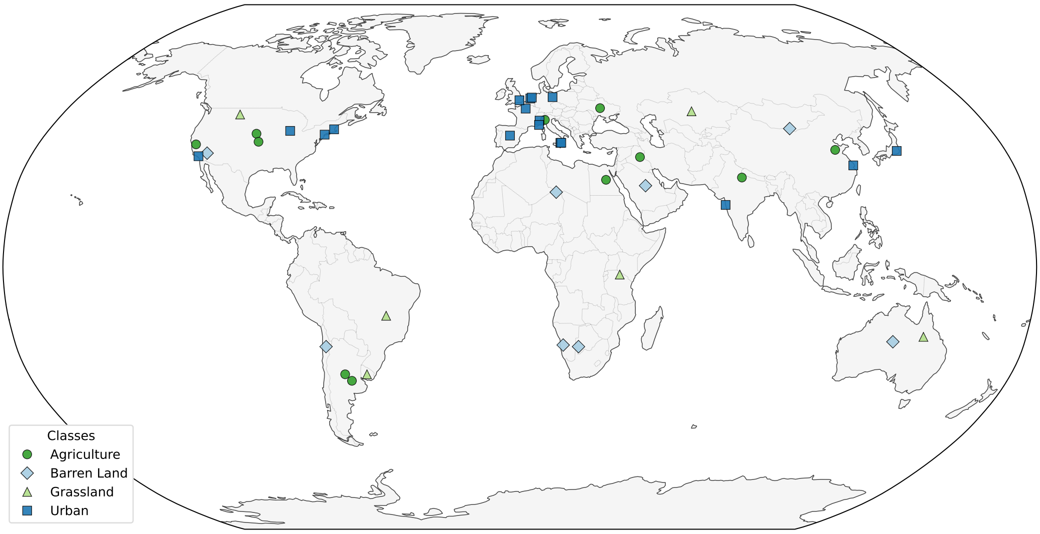


Fig. 15. Highlighted points indicate the Areas of Interest (AOIs) used for end-to-end evaluation. Colors represent the different classified biomes: Agriculture (●), Barrenland (◆), Grassland (▲), and Urban (■)

Table 3. Detailed geographic coordinates and biome classification of the Areas of Interest (AOIs) in the *STAR-Bench* dataset.

Biome	**Location Name**	**Latitude**	**Longitude**
● Agricultural	Argentine Plains, Córdoba	-33.00	-64.00
	Central Valley, California	37.70	-120.99
	Fertile Crescent, Middle East	33.80	44.20
	Great Plains, North America	41.00	-100.00
	Indo-Gangetic Plain	27.50	80.00
	Kansas Plains, USA	38.50	-98.00
	Nile Valley, Egypt	26.80	31.00
	North China Plain	36.00	117.00
	Pampas, South America	-35.00	-62.00
	Po Valley, Italy	45.31	9.78
	Ukrainian Plains	49.00	32.00
■ Barrenland	Arabian Desert, Asia	25.00	45.00
	Atacama Desert, South America	-24.50	-69.25
	Australian Outback	-23.00	133.00
	Gobi Desert, Asia	42.59	103.43
	Kalahari Desert, Africa	-24.50	21.00
	Mojave Desert, USA	35.00	-115.50

Table 3 – Continued from previous page

Biome	Location Name	Latitude	Longitude
	Namib Desert, Africa	-24.00	15.50
	Sahara Desert, Africa	23.00	13.00
▲ Grassland	Cerrado Savanna, Brazil	-15.00	-47.00
	Downs, Australia	-21.50	143.50
	Eurasian Steppe	48.00	68.00
	Montana Prairie, USA	47.00	-110.00
	Pampas, Uruguay	-33.00	-56.00
	Serengeti, Africa	-2.33	34.83
♦ Urban	Amsterdam, Netherlands	52.37	4.90
	Barcellona Pozzo di Gotto, Italy	38.16	15.21
	Berlin, Germany	52.52	13.41
	Boston, USA	42.36	-71.06
	Chicago, USA	41.88	-87.63
	London, UK	51.51	-0.13
	Los Angeles, USA	34.05	-118.24
	Madrid, Spain	40.42	-3.70
	Messina, Italy	38.19	15.55
	Monaco	43.74	7.42
	Mumbai, India	19.08	72.88
	New York City, USA	40.71	-74.01
	Paris, France	48.86	2.35
	Shanghai, China	31.23	121.47
	The Hague, Netherlands	52.08	4.29
	Tokyo, Japan	35.68	139.65
	Turin, Italy	45.07	7.69

A.1.1 Sampling Logic. To capture the temporal and radiometric variability inherent in satellite observations, each location on the map was sampled at 3-month intervals over a full annual cycle for 3 years.

A.2 Framework Abstraction Layer

This section provides the implementation details of the core abstraction layer within the *STAR-Bench* framework. To ensure a consistent evaluation environment across diverse hardware backends, we enforce a standardized interface through the BaseRegisterer abstract class.

As shown in Listing 1, the framework encapsulates the boilerplate logic. Automatically detecting and managing the target device (CPU, CUDA-enabled GPU, or FPGA). Standardizing the tensor conversion process to ensure that image data is correctly resident in the memory space accessible by the hardware accelerators (e.g., handling memory-mapped I/O for the FPGAManager). Separating the I/O management and pre-processing from the core registration algorithm, which is implemented in the _do_registration method.

```
import torch
import cv2

class BaseRegisterer(ABC):
```

```
    """Abstract Base Class for standardized registration methods."""
    def __init__(self, device: str):
        # Hardware-aware device selection
        if device == 'cuda' and torch.cuda.is_available():
            target = 'cuda'
        elif device == 'fpga':
            target = 'cpu'
            self.fpga_manager = FPGAManager()
        else:
            target = 'cpu'

        self.device = torch.device(target)
        logger.info(f"Initializing␣{self.name}␣on␣{self.device}")

    def register(self, ref_path: str, flt_path: str) -> dict:
        # Boilerplate for loading and tensorization
        ref_img = cv2.imread(ref_path, cv2.IMREAD_GRAYSCALE)
        flt_img = cv2.imread(flt_path, cv2.IMREAD_GRAYSCALE)

        # Move data to the selected device memory (CPU or GPU)
        ref_tensor = torch.as_tensor(ref_img, device=self.device)
        flt_tensor = torch.as_tensor(flt_img, device=self.device)

        return self._do_registration(ref_tensor, flt_tensor)

    @abstractmethod
    def _do_registration(self, ref: torch.Tensor, flt: torch.Tensor) -> dict:
        """Specific implementation for XFeat, SIFT, or FPGA-Native paths."""
        pass
```

Listing 1. Standardized device management and image-to-tensor conversion within the *STAR-Bench* framework.

B Detailed Experimental Characterization

This appendix provides an exhaustive statistical characterization of the results discussed in the main text. The analysis is divided into two parts: **Baseline Results (*STAR-Bench*)**, which defines the performance limits of standard pipelines, and ***ETNA* Results**, which illustrates the performance envelope of our proposed hardware-accelerated framework.

The following symbols are used throughout:

- **Modality**: Intra-band (①), Monomodal (②), Multimodal (③), and Multi-Polarity (④).
- **Terrain**: Agriculture (ⓐ), Barrenland (ⓑ), Grassland (ⓒ), and Urban (ⓓ).
- **Deformation**: Nominal (I), Off-Nominal (II), and Anomaly (III).

B.1 Utilization of the IR methods

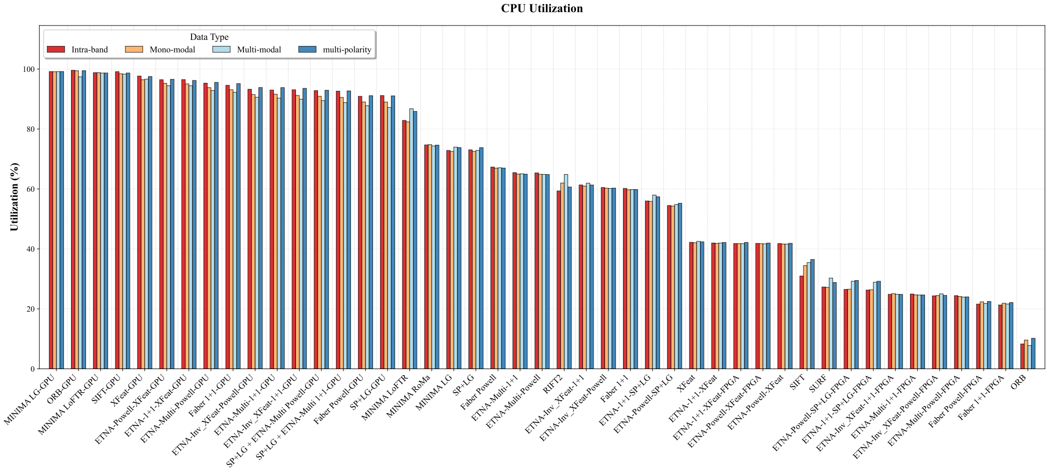
B.2 FPGA Resource Utilization

Table 4. Resource utilization for the *ETNA* hardware module on the Kria K26 SOM.

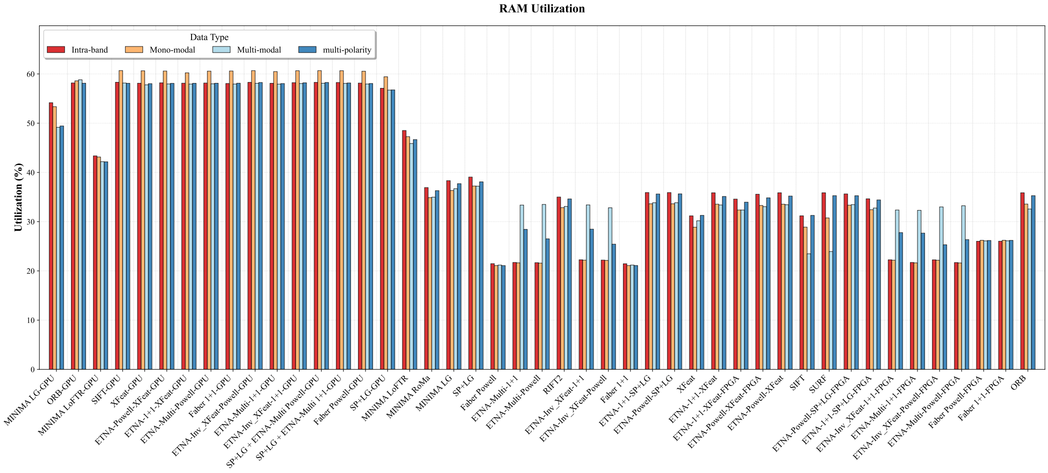
	Used	Available	Utilization (%)
LUTs	25,390	117,120	21.68
FFs	27,015	234,240	11.53
BRAM	68.5	144	47.57
DSPs	102	1,248	8.17

B.3 Detailed Results of *STAR-Bench* Characterization

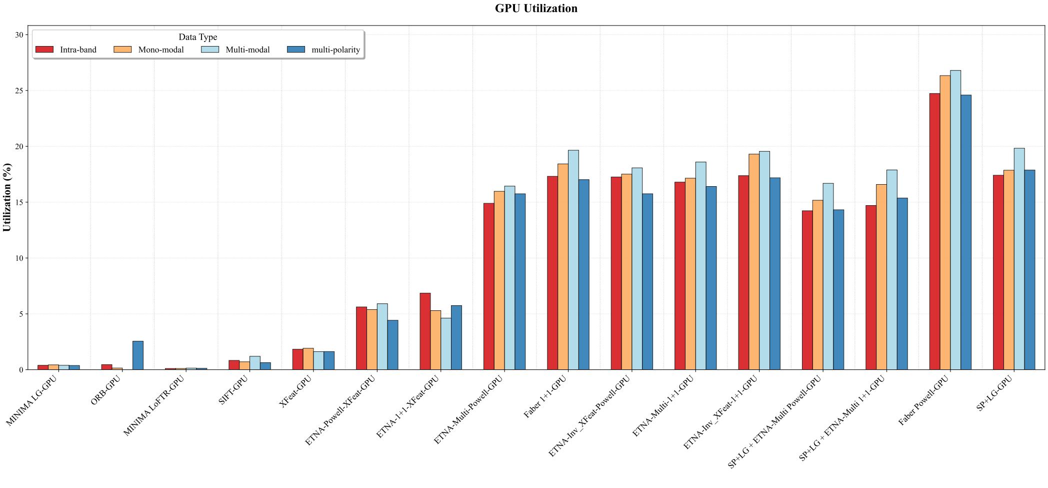
B.3.1 STAR-Bench- Analysis Across Sensing Modalities. This section extends the discussion in §4.1 by providing the full Pareto characterization for baseline pipelines across all four sensing modalities. We report median, P95, and P99 metrics to quantify the reliability gap between average performance and worst-case operational scenarios.



(a) CPU Utilization

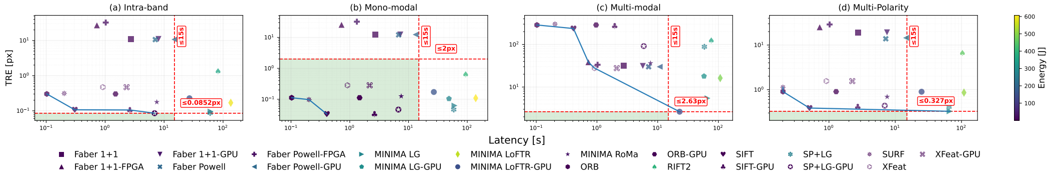


(b) RAM Utilization

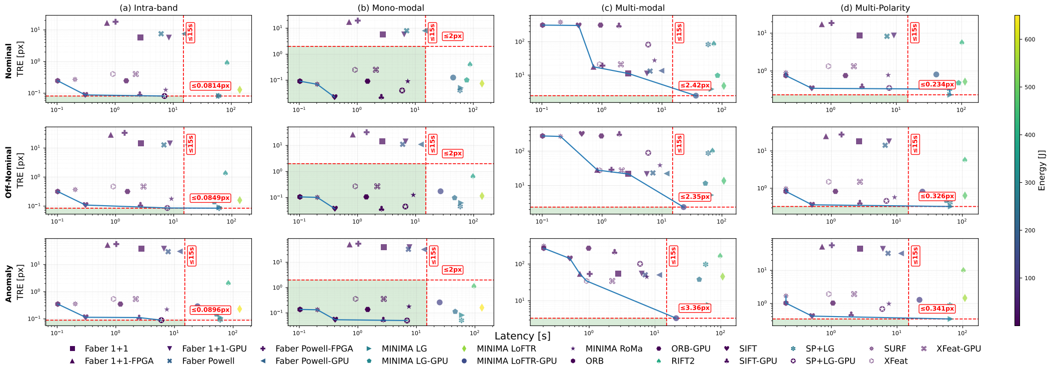


(c) GPU Utilization

Fig. 16. Baseline Pareto characterization with **Median** aggregator.

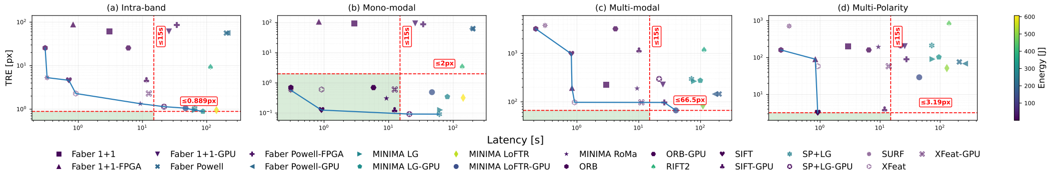


(a) Median aggregator summary (1x4 view).

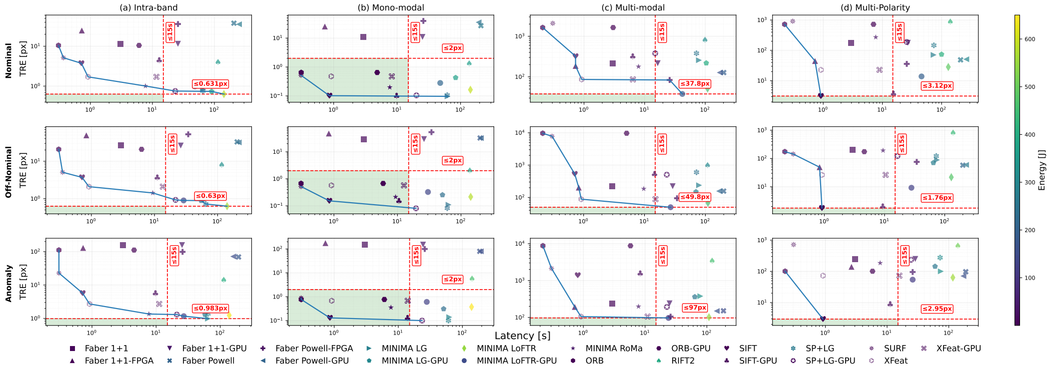


(b) Detailed Median characterization (3x4 view).

Fig. 17. Baseline Pareto characterization with **Median** aggregator.

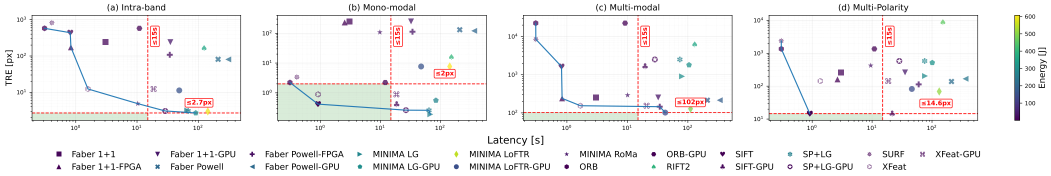


(a) P95 aggregator summary (1x4 view).

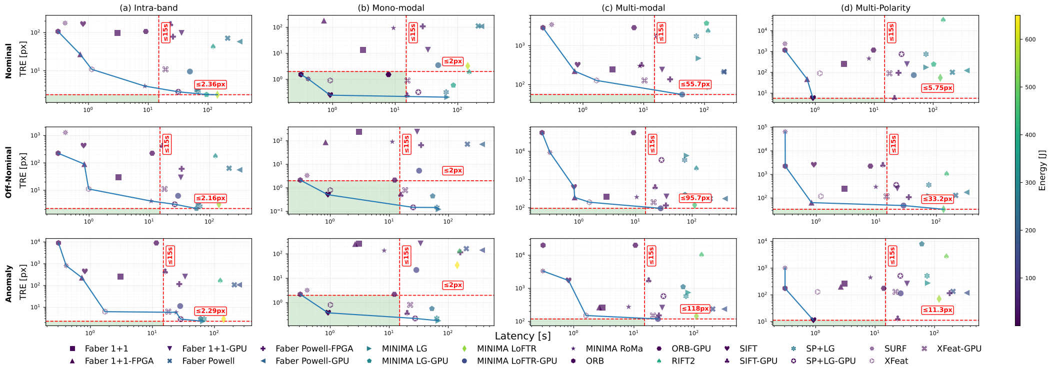


(b) Detailed P95 characterization (3x4 view).

Fig. 18. Baseline Pareto characterization with **P95** aggregator.



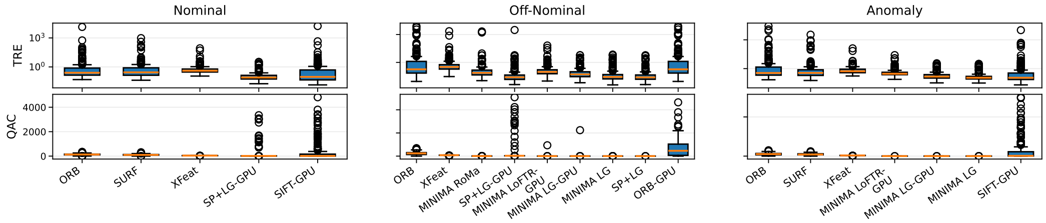
(a) P99 aggregator summary (1x4 view).



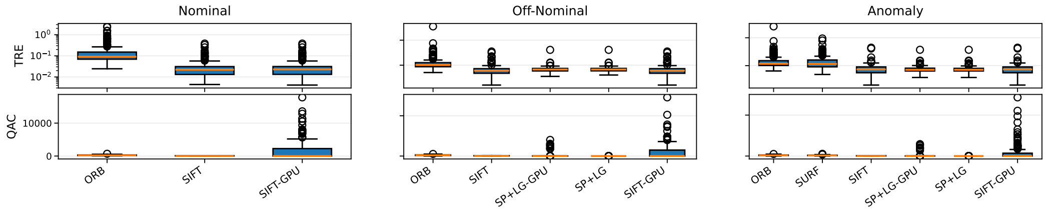
(b) Detailed P99 characterization (3x4 view).

Fig. 19. Baseline Pareto characterization with **P99** aggregator.

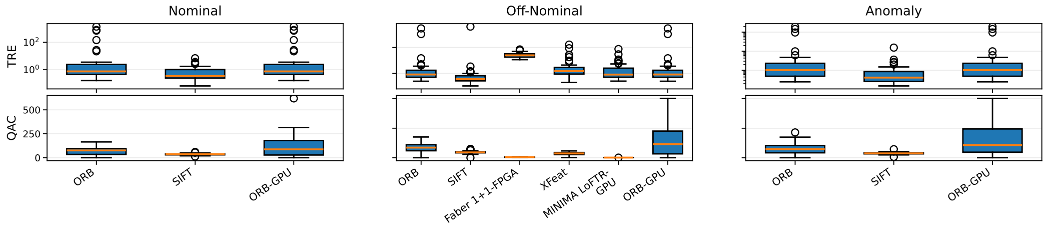
B.3.2 STAR-Bench- Parallax Stability. The following analysis complements §4.2 by illustrating how geometric stress (Ⅰ, Ⅱ, Ⅲ) affects the stability of the baseline descriptors. These plots highlight the rapid degradation of TRE as initial misalignments exceed nominal conditions.



(a) Intra-band (①) baseline stability.



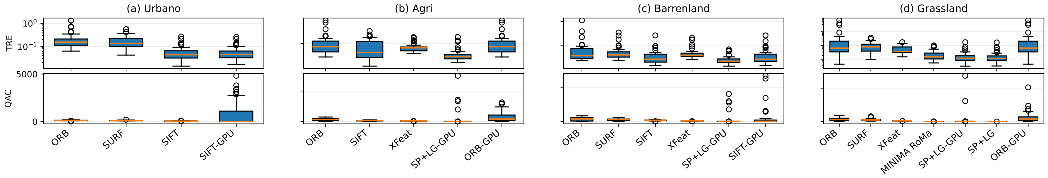
(b) Monomodal (②) baseline stability.



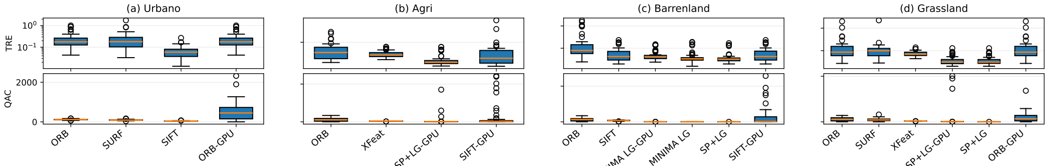
(c) Multi-Polarity (④) baseline stability.

Fig. 20. Baseline geometric resilience across sensing modalities and operational tiers.

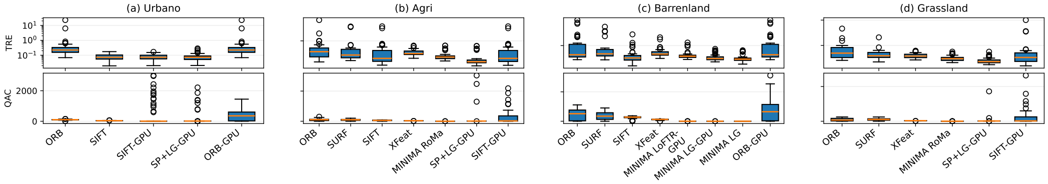
B.3.3 STAR-Bench- Analysis Across Terrain Variability. As discussed in §4.3, textural features of the landscape significantly impact registration success. This section breaks down the baseline error distribution across the four biomes (ⓐ - ⓓ), identifying Barrenland as the most critical failure point for standard methods.



(a) Nominal (I)

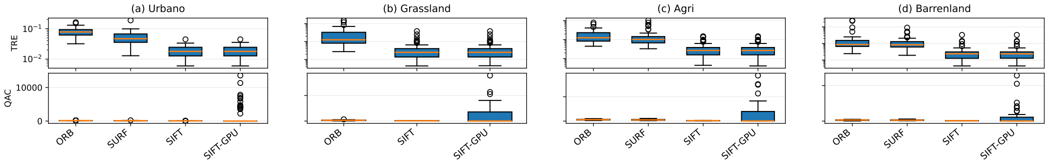


(b) Off-Nominal (II)

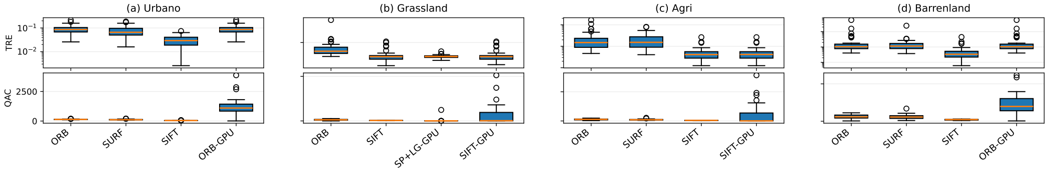


(c) Anomaly (III)

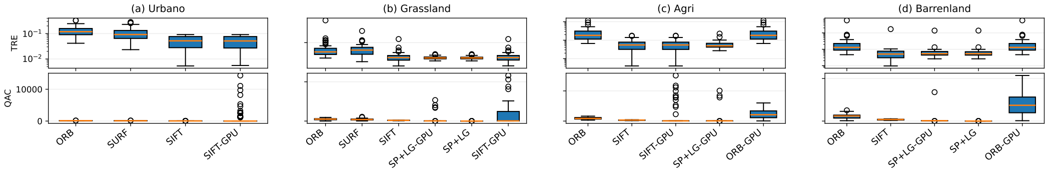
Fig. 21. Baseline distributions for **Intra-band** (①) sensing across all deformation tiers.



(a) Nominal (I)

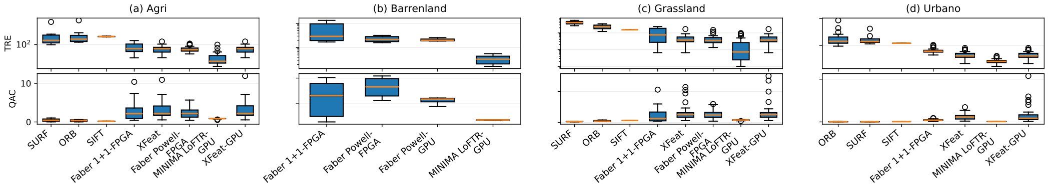


(b) Off-Nominal (II)

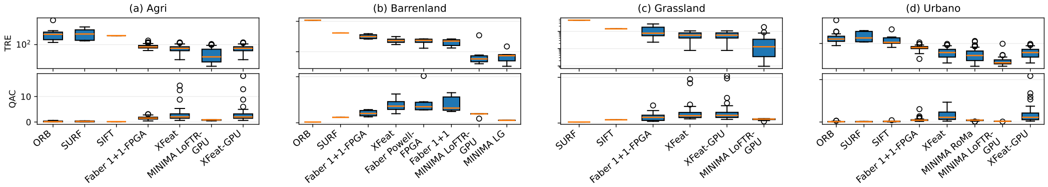


(c) Anomaly (III)

Fig. 22. Baseline distributions for **Mono-modal** (2) sensing across all deformation tiers.

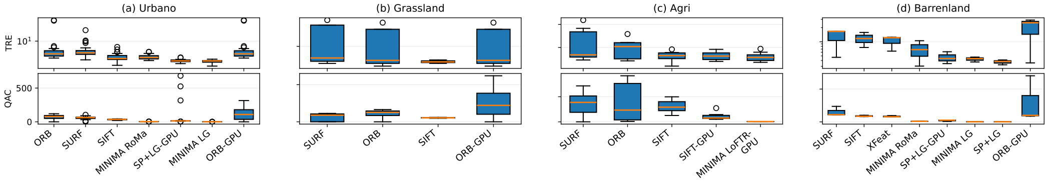


(a) Off-Nominal (II)

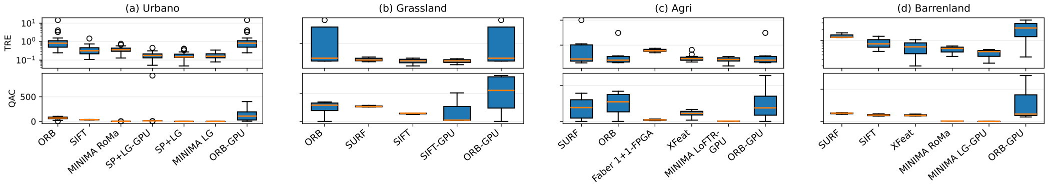


(b) Anomaly (III)

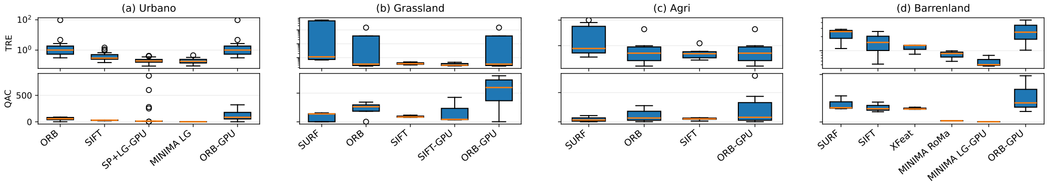
Fig. 23. Baseline distributions for **Multi-modal** (3) sensing across all deformation tiers.



(a) Nominal (I)



(b) Off-Nominal (II)

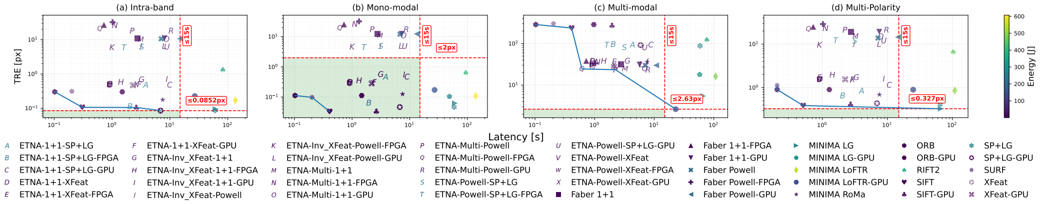


(c) Anomaly (III)

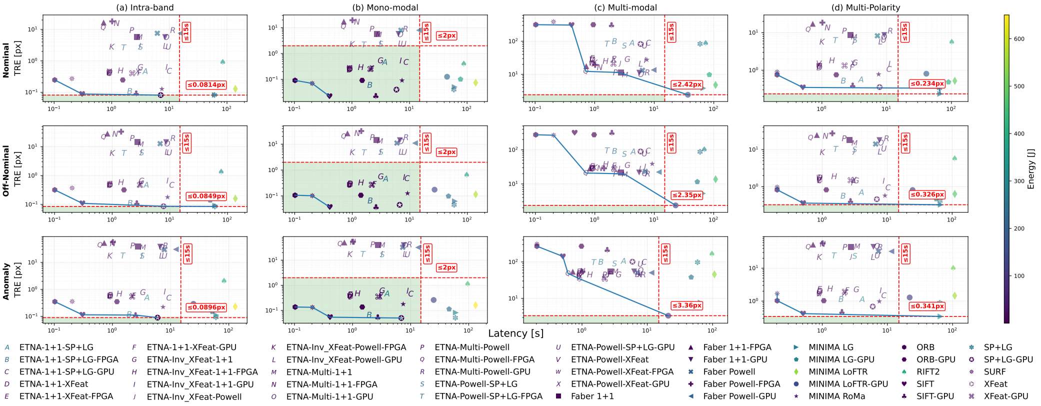
Fig. 24. Baseline distributions for **Multi-Polarity** (④) sensing across all deformation tiers.

B.4 Detailed Results of *ETNA*

B.4.1 ETNA- Analysis Across Sensing Modalities. This section details the performance gains achieved by the *ETNA* framework, as introduced in §6.1. The use of hardware-accelerated intensity engines enables *ETNA* to maintain tight P99 error bounds even in modalities that typically cause functional collapse in baseline pipelines.

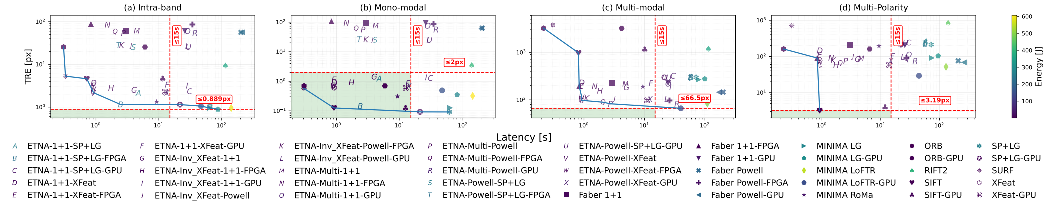


(a) Median aggregator summary (1x4 view).

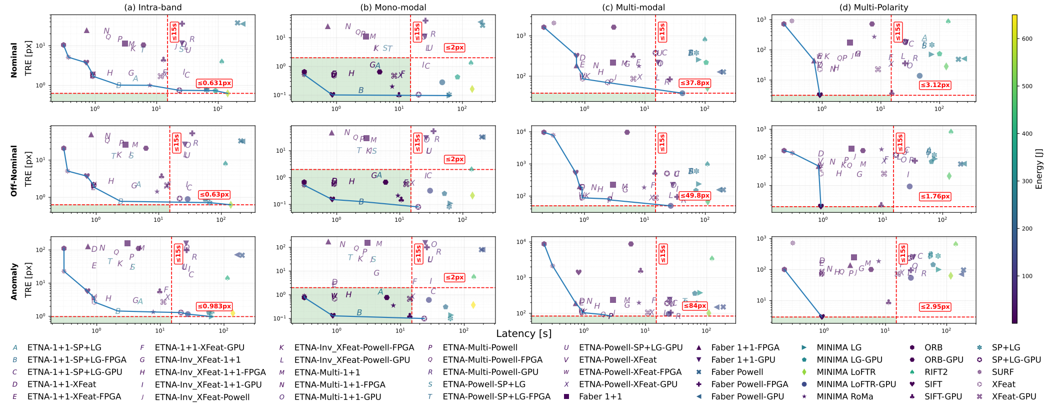


(b) Detailed Median characterization (3x4 view).

Fig. 25. *ETNA* performance across sensing modalities with **Median** aggregator.

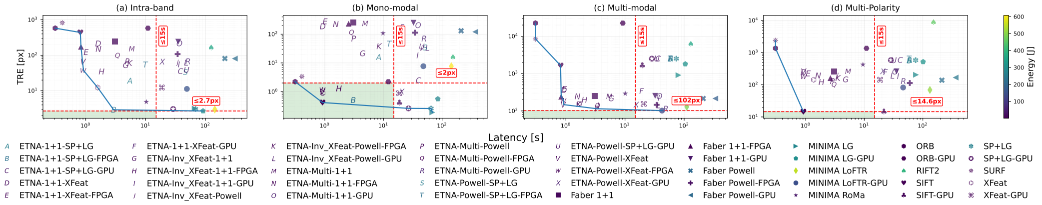


(a) P95 aggregator summary (1x4 view).

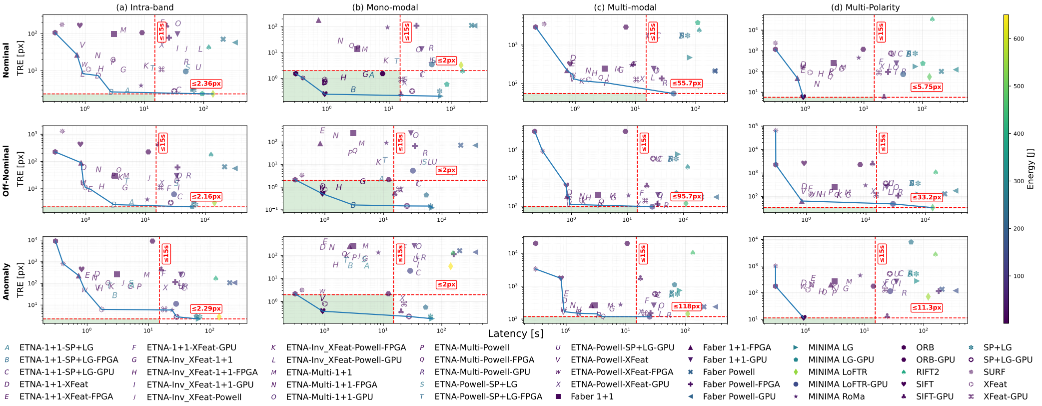


(b) Detailed P95 characterization (3x4 view).

Fig. 26. *ETNA* performance across sensing modalities with **P95** aggregator.



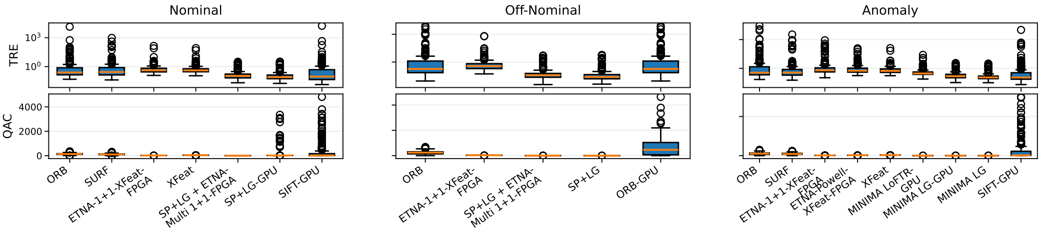
(a) P99 aggregator summary (1x4 view).



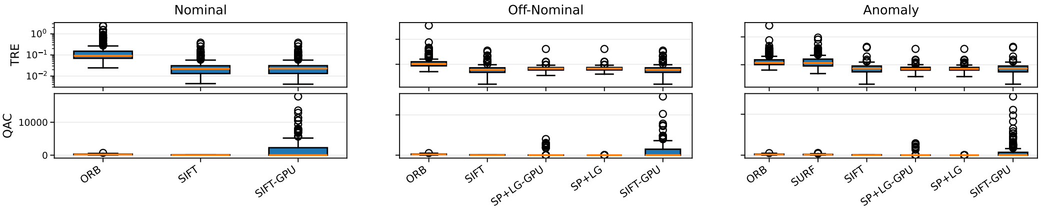
(b) Detailed P99 characterization (3x4 view).

Fig. 27. *ETNA* performance across sensing modalities with **P99** aggregator.

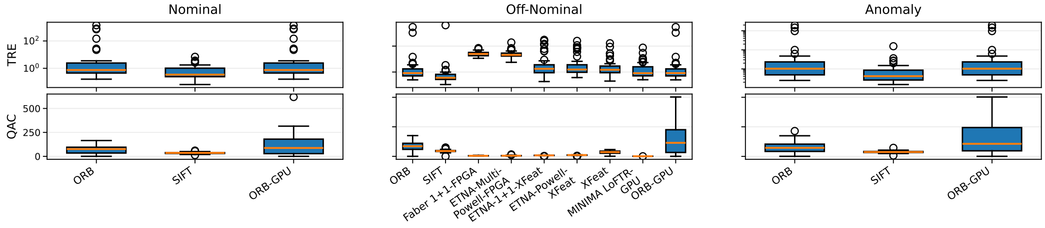
B.4.2 ETNA- Parallax Stability. As discussed in §6.2, the robustness of *ETNA* is particularly evident under increasing geometric stress. These plots demonstrate that our framework maintains a sub-pixel TRE even in Anomaly (III) tiers, whereas baselines exhibit catastrophic errors.



(a) Intra-band (①) resilience.



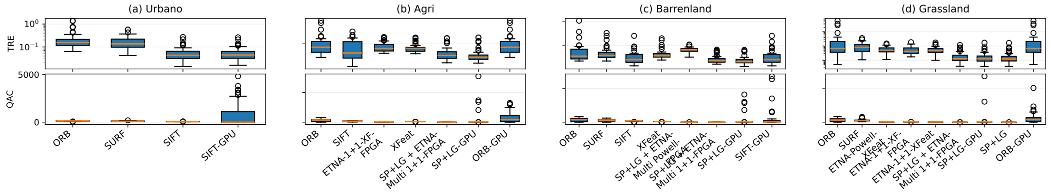
(b) Monomodal (②) resilience.



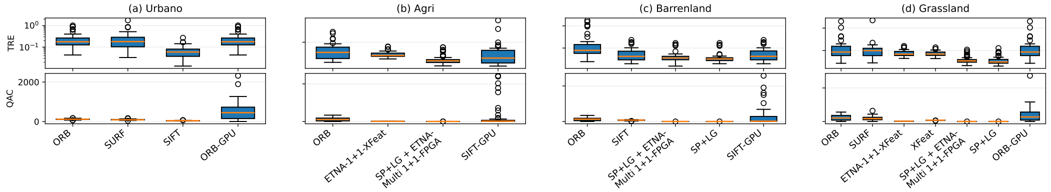
(c) Multi-Polarity (④) resilience.

Fig. 28. Geometric resilience across operational tiers (I, II, III) for *ETNA*.

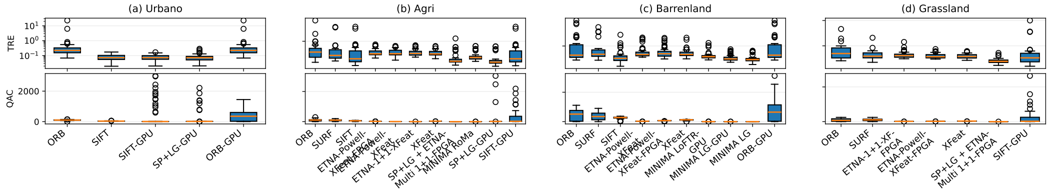
B.4.3 ETNA- Analysis Across Terrain Variability. Final breakdown of *ETNA* stability across Agriculture (ⓐ), Barrenland (ⓑ), Grassland (ⓒ), and Urban (ⓓ), complementing the analysis in §6.3. The results demonstrate that *ETNA* successfully bridges the "textural gap" that affects point-based baselines.



(a) Nominal (I)



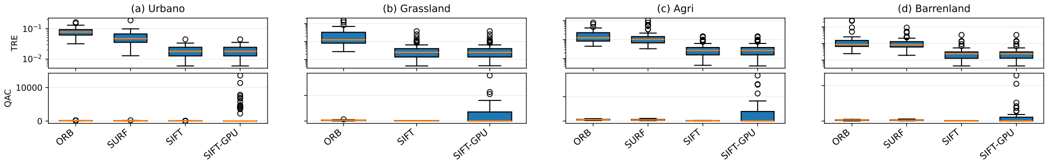
(b) Off-Nominal (II)



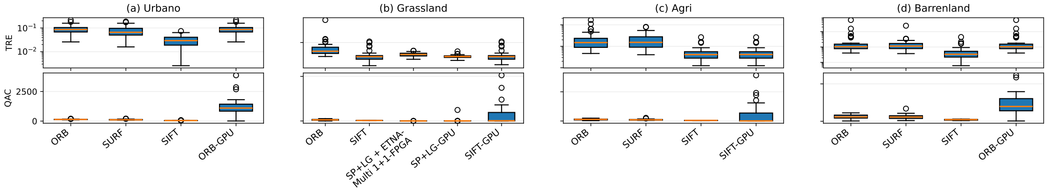
(c) Anomaly (III)

Fig. 29. *ETNA* stability across biomes for **Intra-band** (①) sensing (ETNA dataset).

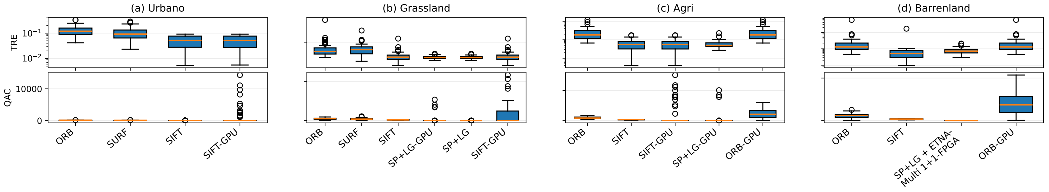
Received January 2026; revised March 2026; accepted April 2026



(a) Nominal (I)

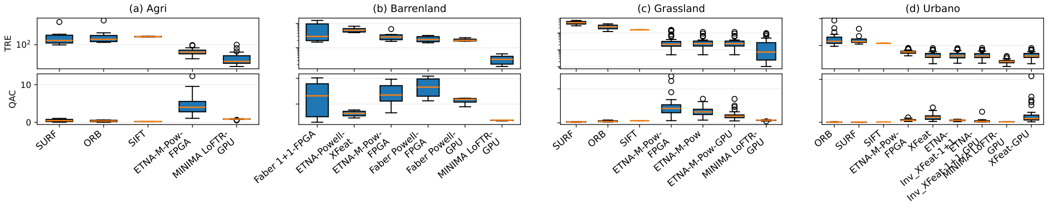


(b) Off-Nominal (II)

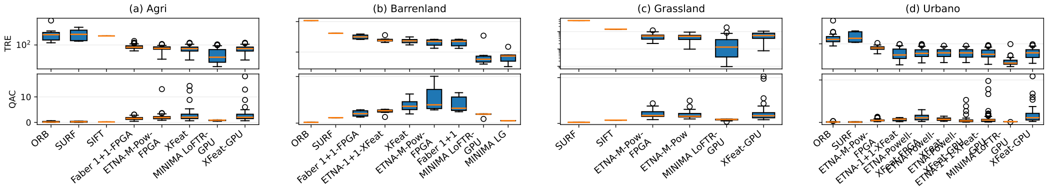


(c) Anomaly (III)

Fig. 30. *ETNA* stability across biomes for **Mono-modal** (②) sensing (ETNA dataset).

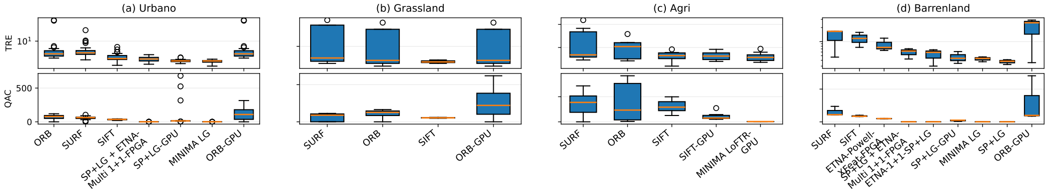


(a) Off-Nominal (II)

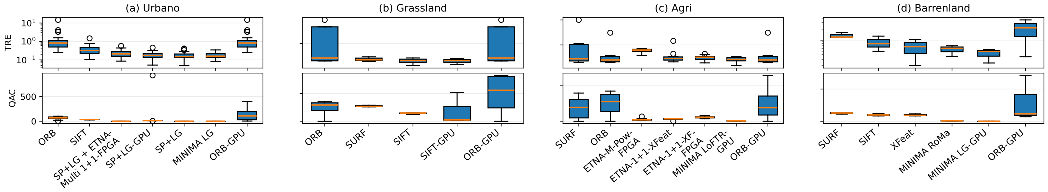


(b) Anomaly (III)

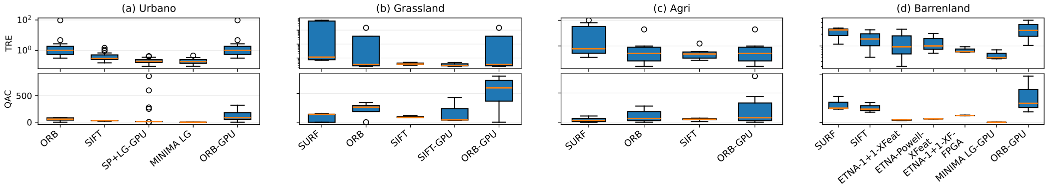
Fig. 31. *ETNA* stability across biomes for **Multi-modal** (③) sensing (ETNA dataset).



(a) Nominal (I)



(b) Off-Nominal (II)



(c) Anomaly (III)

Fig. 32. *ETNA* stability across biomes for **Multi-Polarity** (4) sensing (ETNA dataset).