ASTRA: A MULTI-SATELLITE BENCHMARK FOR ONBOARD 6-DOF SPACECRAFT POSE ESTIMATION

Anonymous authors
Paper under double-blind review



Figure 1: The twenty NASA-referenced spacecraft included in ASTRA, covering crewed, Earth-observation, heliophysics, planetary, and space-telescope missions across four decades of mission design and an order of magnitude in physical scale.

ABSTRACT

Accurate six-degree-of-freedom (6-DoF) pose estimation of non-cooperative satellites is critical for on-orbit servicing, assembly, and manufacturing missions, yet progress is bottlenecked by a severe scarcity of labeled in-orbit imagery and by the narrow target coverage of existing public datasets, most of which focus on only one or two spacecraft. We introduce ASTRA, a large-scale multi-satellite benchmark for onboard 6-DoF pose estimation that spans twenty structurally diverse targets drawn from NASA reference CAD models. ASTRA pairs every target with two complementary data regimes under a common annotation format, comprising an independent-frame training subset rendered under broad parameter sweeps for diverse pose coverage and a sequential test subset of rendezvous trajectories. To close the sim-to-real gap without relying on physical hardware-in-the-loop (HIL) facilities, ASTRA is released with two orthogonal real-style generation pipelines. FRESCO is a neural operator that performs band-selective Fourier amplitude mapping while preserving pose-critical geometry, and a world-model pipeline based on NVIDIA COSMOS turns sequential edge maps from synthetic renderings into photorealistic frames with diverse space environments.

1 INTRODUCTION

Robust 6-DoF pose estimation Kendall et al. (2015); Chen et al. (2022) of a target spacecraft is a prerequisite for autonomous proximity operations, on-orbit servicing, and active debris removal. Progress has long been held back by two intertwined obstacles. Public benchmarks are confined to only one or two spacecraft, with SPEED Kisantal et al. (2020), SPEED+ Park et al. (2022), and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

SHIRT Park & D'Amico (2023) all built around the Tango spacecraft, DLVS3 Velkei et al. (2025) centered on the Hubble Space Telescope, and only SPARK Musallam et al. (2021) extending beyond a single target. Reported numbers therefore reflect single-target overfitting rather than operationally relevant generalization across the geometric and photometric diversity of real spacecraft. Labeled in-orbit imagery is likewise essentially unavailable, forcing researchers to rely on physical hardware-in-the-loop (HIL) facilities Park et al. (2021); Gill et al. (2007), which are expensive, low-throughput, and tied to a single mockup, leaving them non-scalable across spacecraft geometries.

We introduce ASTRA, a large-scale multi-satellite benchmark for 6-DoF pose estimation that addresses both issues jointly. ASTRA spans twenty structurally diverse spacecraft drawn from the publicly available NASA 3D Resources repository[1], covering crewed vehicles, Earth-observation buses, heliophysics and planetary probes, and large space telescopes. Every target is rendered through a unified Unreal Engine 5 (UE5) pipeline[2] that imports each NASA CAD model with its native materials and articulates rigid subcomponents where applicable. Although fully synthetic, the scenes are designed to mirror real space imagery, with Earth as the dominant background and direct sunlight as the primary light source, as illustrated for the Rosetta spacecraft in Figure 2. Every frame carries a full 6-DoF pose and up to 39 visible 3D keypoints, captured at one of two camera fields-of-view under a single common format shared across all twenty targets.

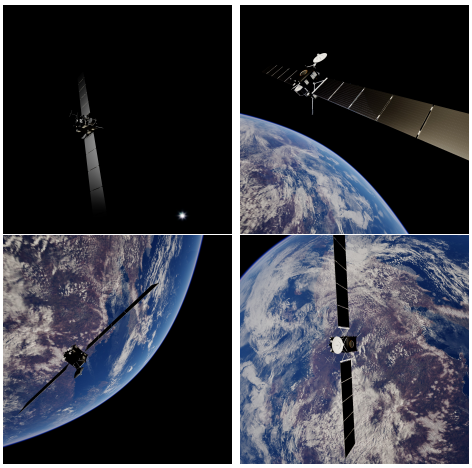


Figure 2: UE5-rendered Rosetta spacecraft with realistic various backgrounds.

To eliminate dependence on HIL facilities and expose the sim-to-real gap controllably, ASTRA further provides two complementary real-style generation pipelines that apply uniformly to every target. FRESCO (FREquency-SpaCe neural Operator) is a learned Fourier-domain operator that modifies the amplitude spectrum while preserving phase, so that the underlying geometric structure is retained as realistic illumination and texture statistics are introduced. The second pipeline conditions NVIDIA COSMOS Agarwal et al. (2025) on high-fidelity edge maps derived from the UE5 renderings to synthesize photorealistic frames with diverse orbital backgrounds. FRESCO preserves geometry by construction, whereas COSMOS introduces broader appearance and background variation that complements FRESCO's geometry-preserving spectral mapping. We additionally benchmark representative pose estimation models on the NVIDIA Jetson series for onboard inference.

This work makes three contributions:

- **ASTRA benchmark.** We release a unified multi-satellite benchmark of twenty NASA-referenced spacecraft generated through a single UE5+NASA-CAD pipeline, to our knowledge the largest and most target-diverse public benchmark for spacecraft pose estimation.
- **Dual sim-to-real pipelines.** We release FRESCO, a learned Fourier-domain operator, together with a COSMOS-based world-model pipeline conditioned on UE5 edge maps, providing two complementary sources of real-style imagery for every target.
- **HIL-free protocol.** We define a standardized protocol that trains on synthetic imagery and evaluates on FRESCO- or COSMOS-generated frames, eliminating reliance on physical HIL facilities and applying uniformly to all twenty targets.

2 Related Work

Spacecraft Pose Estimation Datasets. The lack of labeled in-orbit imagery has motivated a series of synthetic and HIL datasets. SPEED Kisantal et al. (2020) introduced the Tango spacecraft from the PRISMA mission Gill et al. (2007) with synthetic renderings paired with a small set of

[1] https://nasa3d.arc.nasa.gov/

[2] https://www.unrealengine.com/

Table 1: Comparison of public satellite pose estimation datasets. ASTRA combines multi-target coverage, sequential trajectories, and two sim-to-real pipelines.

Dataset Name	# Target	Total Images	Sim2Real Method	Sequential	#FOV	Keypoints	Resolution
SPEED Kisantal et al. (2020)	1 (Tango)	∼15k	TRON	No	1	Yes	1920×1200
SPEED+ Park et al. (2022)	1 (Tango)	∼70k	TRON	No	1	Yes	1920×1200
SHIRT Park & D'Amico (2023)	1 (Tango)	∼5k	TRON	Yes	1	Yes	1920×1200
SPARK (stream 1) Musallam et al. (2021)	11	∼150k	N/A	No	1	No	1440×1080
SPARK (stream 2) Musallam et al. (2021)	1 (Proba-2)	∼150k	N/A	Yes	1	No	1440×1080
DLVS3 Velkei et al. (2025)	1 (HST)	∼1,000k	N/A	Yes	1	No	1024×1024
ASTRA (Ours)	**20**	**∼1,576k**	**FRESCO & COSMOS**	**Yes**	**2**	**Yes**	**2048×2048**

TRON-captured images, and SPEED+ Park et al. (2022) extended realism by adding HIL samples under two calibrated illumination domains, forming the basis of the Satellite Pose Estimation Challenge (SPEC2021) Park et al. (2023). SHIRT Park & D'Amico (2023) introduced sequential imagery through paired synthetic and TRON-captured trajectories Park & D'Amico (2024a), while DLVS3 Velkei et al. (2025) provided a fully synthetic large-scale dataset; both, however, remain restricted to a single target. SPARK Musallam et al. (2021) broadened object diversity to eleven spacecraft Park & D'Amico (2024), but offers limited variation per target and no sim-to-real coverage. As summarized in Table 1, prior datasets each emphasize a single axis, whereas ASTRA jointly delivers target diversity, sequential trajectories, and paired real-style imagery on every target. This breadth also enables, for the first time at this scope, the study of RNN-based trackers Xu et al. (2022), flow-assisted filtering Piga et al. (2021); Krishnan et al. (2015), and test-time adaptation Lee et al. (2023); Tian et al. (2024) across multiple spacecraft.

Bridging the Sim-to-Real Gap. Traditional approaches to the sim-to-real gap operate in pixel or feature space through domain randomization Tobin et al. (2017) and adversarial alignment Ganin & Lempitsky (2015); Tzeng et al. (2017), or in the frequency domain Yang & Soatto (2020); Xu et al. (2021); Huang et al. (2021), exploiting the amplitude–phase disentanglement of the Fourier transform to transfer global appearance while preserving the structural cues encoded in phase. Existing frequency-domain methods such as Fourier Domain Adaptation (FDA) Yang & Soatto (2020) are training-free and apply fixed amplitude swaps between domains, whereas FRESCO learns a Fourier-domain operator directly from paired synthetic–real data (Appendix D). More recently, conditional generative world models such as NVIDIA COSMOS Agarwal et al. (2025) bridge this gap by synthesizing photorealistic scenes from a text prompt while a structural conditioning signal such as an edge map anchors the underlying geometry Zhang et al. (2023). We adopt this approach through a COSMOS-based pipeline, detailed together with FRESCO in Section 4.

Onboard Computing for Satellites. Onboard processing on Earth-observation and small satellites is increasingly used to reduce downlink volume and to enable closed-loop autonomy. Reported applications include onboard cloud detection for downlink filtering Giuffrida et al. (2020), methane and trace-gas detection from hyperspectral imagery Růžička et al. (2023); Herec et al. (2025), and image compression to reduce raw payload volume before downlink Guerrisi et al. (2023); Cho et al. (2026); Park et al. (2026). These workloads share a common constraint: they must run within tight latency, memory, and power envelopes on embedded accelerators. NVIDIA's Jetson family Mittal (2019); Karumbunathan (2022) is the standard reference platform for this regime, providing a range of compute and power envelopes that approximate those of small-satellite payloads. Spacecraft pose estimation belongs to the same onboard workload mix, since proximity-operations applications such as on-orbit servicing, docking, and active debris removal require the chaser to estimate the target's pose locally at frame rate rather than relaying imagery to ground.

3 ASTRA: Autonomous Spacecraft Tracking and Rendezvous Archive

ASTRA is a multi-satellite benchmark dataset composed of per-target datasets that share a common rendering pipeline, scenario family, and annotation format. The dataset is generated end-to-end by a single Unreal Engine 5 (UE5) simulator built around publicly available NASA reference CAD models, so that the same evaluation protocol applies uniformly to every spacecraft.

3.1 TWENTY NASA-REFERENCED SPACECRAFT

ASTRA covers twenty NASA-referenced spacecrafts selected to maximize geometric and mission-domain diversity within publicly redistributable assets. Figure 3 provides a visual overview, with targets ranging from compact crewed vehicles, such as Gemini and Apollo–Soyuz, to large space telescopes, such as Hubble and JWST. The collection spans roughly four decades of mission design, from early 1960s crewed capsules to next-generation observatories, and an order of magnitude in physical scale, exposing a wide variety of articulated structures such as solar arrays, antennae, sensor booms, and sunshields. The targets also exhibit substantially different specular and diffuse material distributions, since each generation of spacecraft adopts different thermal blankets, optical coatings, and surface treatments. Together these factors produce a substantially more demanding generalization regime than any prior single-target benchmark. The full list with mission category, agency, orbital regime, year, and primary mission is provided in Appendix B.



Figure 3: The 20 spacecraft covered by ASTRA.

3.2 RENDERING PIPELINE AND ANNOTATIONS

All twenty targets are rendered with a single UE5 simulator that ingests each NASA CAD model with its native materials, articulates rigid sub-components where the source asset provides them, and places the target inside a virtual environment with Earth-albedo and direct-sunlight illumination. A lightweight orbital-dynamics module drives the relative motion between a chaser camera and the target, and we additionally enable in-engine post-processing effects such as lens flare, blooming, vignetting, and motion blur to approximate the photometric signature of real onboard sensors. Each target is captured under two camera fields of view (Figure 10), with a narrow 45° view simulating inspection cameras and a wide 90° view simulating navigation cameras, and the chaser-to-target distance envelope is set per target as a function of the spacecraft's bounding-sphere radius so that the target consistently occupies a meaningful fraction of the image plane.

Table 2: ASTRA rendering protocol, applied uniformly to all spacecraft.

Property	Value
# Training data	1,000k
# Test data	576k
Resolution	2048×2048
Frame rate for test	24 fps
Frames per trajectory	240
Camera FOV	45°, 90°
Test scenarios	ROE1, ROE2
File formats	JPG, JSON
Annotation types	6-DoF, 3D keypoints
Total dataset size	$\sim$680GB

Every frame is annotated with the full 6-DoF pose as a quaternion and translation relative to the camera, together with up to 39 visible 3D keypoints that are defined on each CAD model and projected per frame from the known camera matrix, with occluded keypoints flagged using the depth buffer. Annotations follow a unified schema across all twenty targets, supporting cross-target evaluation without format changes. Table 2 summarizes the per-target rendering protocol, and Appendix C describes the full generation pipeline in detail.

3.3 TRAJECTORIES AND DIVERSITY

Following prior methodology for spacecraft rendezvous Park & D'Amico (2023); Koenig et al. (2017), we define two Relative Orbital Element (ROE) trajectory scenarios applied uniformly to every target, with full details in Appendix C. ROE1 is a v-bar hold that maintains a fixed relative position along the target's velocity vector, yielding low visual diversity for controlled evaluation, and ROE2 is an approach scenario in which the chaser gradually closes in on the target. For each target

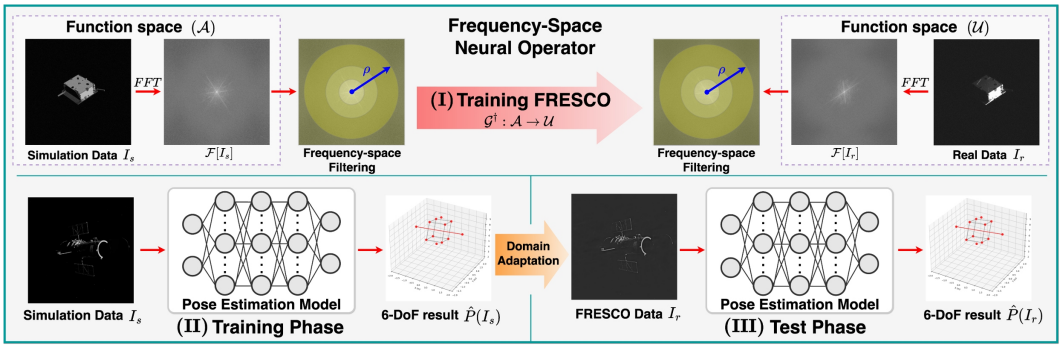


Figure 4: Overall FRESCO pipeline. (I) FRESCO learns Fourier amplitude alignment between synthetic and real imagery; (II) a pose regressor is trained on synthetic images; (III) inference proceeds on FRESCO-transformed images with pose-critical geometry preserved.

and each scenario we render 10 trajectories of 240 frames at both FOVs, together with their FRESCO and COSMOS real-style variants, yielding a sequential test subset of 576k frames. Combined with an independent-frame training subset of 1,000k frames rendered under broad parameter sweeps, this brings the ASTRA scale to roughly 1.58 million annotated frames across twenty spacecraft.

ASTRA combines diversity axes that prior datasets rarely provide together. Target diversity comes from twenty geometrically heterogeneous spacecraft, scenario diversity from the configurations detailed in Appendix C, and appearance diversity from synthetic UE5 renderings paired with two real-style image sets from Section 4. Within each trajectory the simulator captures continuous relative motion, scale changes between far and near field, occlusions from articulated sub-components, and dynamic lighting such as specular glints on rotating solar panels.

4 From Synthetic to Real-Style Imagery

To eliminate dependence on physical HIL facilities, ASTRA includes two real-style generation pipelines that apply to every target. FRESCO offers fine-grained style control while preserving target geometry, and the COSMOS pipeline regenerates scene appearance and background diversity from the same edge conditioning, so the two pipelines together cover both controlled stress tests and broader appearance variation.

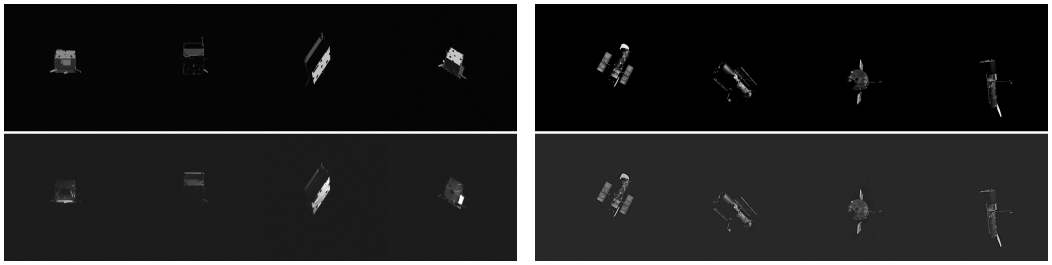
4.1 FRESCO: Neural Operator for Frequency-Space Adaptation

Frequency-domain manipulation is a natural fit for sim-to-real adaptation Yang & Soatto (2020); Xu et al. (2021); Huang et al. (2021), since amplitude carries global appearance while phase encodes geometry. We propose FRESCO, a neural operator that learns a band-selective amplitude filter from synthetic–real image pairs, aligning the synthetic spectrum with the real distribution while keeping the phase fixed (Figure 4). The pose regressor is trained on synthetic images and at test time receives FRESCO-transformed inputs, recovering the same 6-DoF pose under a real-style appearance.

Frequency-space mapping for sim-to-real. FRESCO is built on the Fourier Neural Operator (FNO) framework Li et al. (2020), which parameterizes operators directly in the frequency domain through a learnable spectral filter. Appendix E provides the full background. Let I_s and I_r be a synthetic and a real image defined on $\Omega \subseteq \mathbb{R}^2$, with Fourier transforms $\hat{I}_s = \mathcal{F}[I_s]$ and $\hat{I}_r = \mathcal{F}[I_r]$, and let $\hat{I}(k)$ denote the complex coefficient at frequency mode k. FRESCO learns an operator $\mathcal{O}_\rho$ that reduces the spectral discrepancy between synthetic and real images while preserving the geometric structure encoded in the phase. The operator is parameterized as follows:

$$\hat{I}_{s\to r}(k) = \mathcal{O}_\rho(\hat{I}_s(k)) = R_\varphi(k;\rho) \cdot \hat{I}_s(k), \tag{1}$$

where $R_\varphi(k;\rho)$ is a learnable Fourier filter restricted to a band of radius ρ around the spectral origin. The phase $\angle\hat{I}_{s\to r}(k) = \angle\hat{I}_s(k)$ is left unchanged so that the projected geometry survives



(a) SHIRT: synthetic (top) vs. TRON (bottom). (b) ASTRA: synthetic (top) vs. FRESCO (bottom).

Figure 5: Visual comparison of synthetic (top row) and target-domain (bottom row) appearance. Panel (a) shows SHIRT with paired synthetic and TRON real images, and panel (b) shows HST with synthetic inputs and FRESCO outputs.

the transfer, while low- and mid-frequency amplitudes are modified to inject realistic illumination and texture Oppenheim & Lim (2005). The real-style image is then recovered as follows:

$$I_{s\to r} = \mathcal{F}^{-1}\big[\hat{I}_{s\to r}(k)\big] = \mathcal{F}^{-1}\big[R_\varphi(k;\rho)\cdot\mathcal{F}[I_s](k)\big]. \tag{2}$$

Whereas neural operators are most often applied in physics simulation to learn mappings between function spaces, FRESCO specializes this idea to a mapping between synthetic and real image styles, giving parameterized control over the stylistic distance while leaving target geometry intact.

4.2 COSMOS: World-Model Pipeline

Frequency-domain transformations such as FRESCO preserve geometry by construction, but they cannot introduce new scene content such as diverse backgrounds, lens effects, or alternative illumination styles. To cover this complementary axis of appearance variation while respecting target geometry, ASTRA provides a second pipeline built on NVIDIA COSMOS Agarwal et al. (2025). Among the conditioning modalities supported by COSMOS, such as edge maps and segmentation masks, we choose edge maps as the structural input. COSMOS then generates a photorealistic frame conditioned on the edge map and a text prompt describing the orbital scene and desired style (prompts in Appendix J). Because the edge map encodes the projected geometry of the satellite at a known pose, the generated frame inherits the same pose label as the UE5 render, so no additional labeling is required. Figure 6 illustrates this with Janus composites of Rosetta, Roman, Juno, and Hubble, each placing the UE5 render on the left half and the COSMOS output at the same pose on the right. The satellite outline transitions continuously across the central seam, providing direct visual evidence that the conditional generation preserves the projected pose, while only the right halves exhibit diverse orbital backgrounds together with lens flare, atmospheric limb illumination, and Earth albedo. We distribute both the FRESCO and COSMOS subsets for every target and report results with both in Section 5.

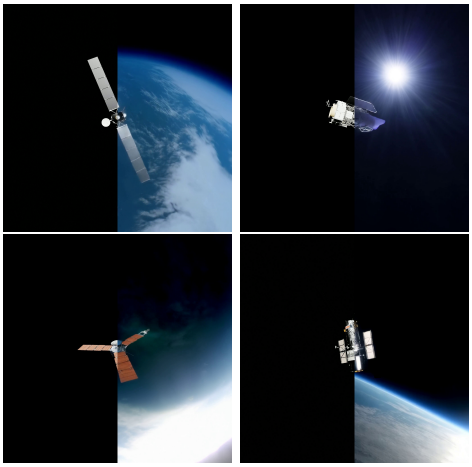


Figure 6: Janus composites of UE5 synthetic (left) and COSMOS-generated (right) frames at identical poses.

5 Experiments

We benchmark five representative pose estimators on ASTRA along three axes, namely accuracy under synthetic, FRESCO, and COSMOS inputs, a controlled domain-interpolation stress test that probes the loss surface between the synthetic and real-style endpoints, and onboard performance

on two Jetson generations under both full-precision and lightweight model variants. All models are trained exclusively on synthetic ASTRA images while evaluation is conducted on synthetic, FRESCO-generated, and COSMOS-generated test images so that no real-style data is observed during training, and we evaluate on SHIRT using models trained on SPEED+ synthetic data. Our training environment is Ubuntu 20.04 LTS with Python 3.12, PyTorch 2.7, and CUDA 12.2 on NVIDIA RTX A6000 GPUs, and onboard evaluation uses NVIDIA JetPack on each Jetson generation with FP32 PyTorch inference. Detailed hyperparameters are provided in Appendix F.

5.1 Baselines

We use five pose estimators that span keypoint regression, multi-task CNNs with heatmap supervision, ViT backbones, and modern real-time detectors, covering both the families that have historically dominated spacecraft pose estimation and recent generic detectors not yet evaluated in this domain. KRN Park et al. (2019) is a lightweight CNN that regresses 2D keypoint locations from which the 6-DoF pose is recovered through a Perspective-n-Point solver Lepetit et al. (2009); Marchand et al. (2015) using the known 3D keypoint coordinates. SPNv2 Park & D'Amico (2024b) is a multi-task CNN that jointly performs bounding-box detection, direct pose regression, heatmap-based keypoint localization, and binary foreground segmentation. SPNv3 Park & D'Amico (2024) adopts a Vision Transformer (ViT) backbone Dosovitskiy et al. (2020) to balance computational efficiency and robustness across the sim-to-real gap. We additionally include two variants of the recent YOLO-26 detector Sapkota et al. (2025), YOLO-26n and YOLO-26s, adapted to predict the same keypoint set used by KRN and combined with the same PnP back-end, which lets us assess how detector-style architectures of different capacity perform on the multi-target ASTRA setting.

5.2 Evaluation Metrics

Let $\hat{q}$ and $\hat{t}$ denote the predicted unit quaternion and translation vector and let q and t denote their ground-truth counterparts expressed in the camera frame. Translation error E_t is the Euclidean distance between the predicted and ground-truth target origins in meters, rotation error E_R is the angular distance between the predicted and ground-truth unit quaternions in degrees, and the unified pose error E_{pose} combines them as

$$E_t = \|\hat{t} - t\|_2, \quad E_R = 2\arccos(|\langle \hat{q}, q \rangle|), \quad E_{pose} = E_R\,[\text{rad}] + \frac{E_t}{\|t\|_2}, \tag{3}$$

where the translation term is normalized by the ground-truth distance $\|t\|_2$ so that position errors are penalized more heavily at close range as required for proximity operations.

5.3 Accuracy under Synthetic, FRESCO, and COSMOS

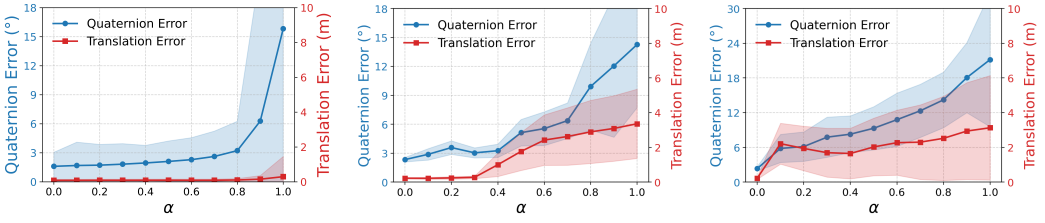
Table 3 reports E_t, E_R, and E_{pose} averaged over the twenty ASTRA targets at both FOVs. On synthetic inputs all five baselines achieve low-degree rotation accuracy with translation error on the order of a few centimeters, confirming that ASTRA renderings are learnable across architectures of widely varying capacity. SPNv2 attains the lowest E_{pose} in every synthetic cell, which we attribute to multi-task supervision supplying geometric cues beyond keypoints. Both real-style domains induce a pronounced accuracy drop in which E_R rises roughly four- to sixfold and E_t rises by an order of magnitude, and COSMOS proves systematically harder than FRESCO especially under the wide 90° FOV where greater background variability further stresses the regressors. KRN degrades the most under domain shift, in line with its reliance on a single keypoint head, whereas SPNv2 retains the smallest gap and remains the strongest baseline in every setting. The two YOLO-26 variants achieve translation accuracy that is competitive with SPNv2 under FRESCO and COSMOS yet exhibit larger rotation errors, indicating that detector-style architectures localize the target reliably while recovering fine-grained orientation less effectively in the multi-target regime.

5.4 Qualitative Comparison and Domain-Interpolation Stress Test

Qualitative domain comparison. Figure 5 contrasts synthetic with real images for SHIRT and synthetic inputs with FRESCO real-style outputs for the HST subset. FRESCO is trained on SHIRT synthetic–real pairs and then applied to HST synthetic frames, and the resulting real-style outputs are

Table 3: Pose estimation performance averaged across all twenty ASTRA targets, under synthetic, FRESCO, and COSMOS input domains at two camera FOV settings. Lower is better for all metrics.

Input Domain	FOV	Metric	KRN	SPNv2	SPNv3	YOLO-26n	YOLO-26s
Synthetic	45°	E_t [m]	0.27 ± 0.21	0.05 ± 0.04	0.10 ± 0.06	0.06 ± 0.04	0.05 ± 0.04
		E_R [°]	4.86 ± 1.60	3.50 ± 1.25	4.48 ± 1.54	5.18 ± 1.71	5.32 ± 1.66
		E_{pose} [-]	0.10 ± 0.03	0.06 ± 0.02	0.08 ± 0.03	0.09 ± 0.03	0.10 ± 0.03
	90°	E_t [m]	0.67 ± 1.31	0.16 ± 0.27	0.34 ± 0.56	0.16 ± 0.30	0.17 ± 0.30
		E_R [°]	3.42 ± 1.18	2.50 ± 0.87	3.87 ± 1.41	3.37 ± 1.18	3.75 ± 1.43
		E_{pose} [-]	0.07 ± 0.03	0.06 ± 0.02	0.08 ± 0.03	0.08 ± 0.03	0.08 ± 0.03
FRESCO	45°	E_t [m]	2.27 ± 3.08	0.57 ± 0.89	0.81 ± 1.26	0.73 ± 1.11	0.42 ± 0.60
		E_R [°]	21.61 ± 10.00	15.00 ± 6.92	18.39 ± 7.23	27.06 ± 10.58	17.56 ± 7.33
		E_{pose} [-]	0.52 ± 0.18	0.29 ± 0.09	0.37 ± 0.13	0.36 ± 0.13	0.45 ± 0.17
	90°	E_t [m]	3.03 ± 3.78	0.65 ± 0.83	1.02 ± 1.24	0.82 ± 0.98	0.59 ± 0.70
		E_R [°]	15.93 ± 14.94	14.50 ± 13.82	19.81 ± 19.02	23.45 ± 21.66	18.47 ± 16.61
		E_{pose} [-]	0.44 ± 0.37	0.29 ± 0.23	0.38 ± 0.35	0.34 ± 0.30	0.48 ± 0.39
COSMOS	45°	E_t [m]	1.86 ± 1.96	0.43 ± 0.46	0.54 ± 0.65	0.39 ± 0.41	0.35 ± 0.41
		E_R [°]	21.28 ± 2.53	19.00 ± 2.25	21.85 ± 2.95	24.89 ± 3.26	24.32 ± 2.90
		E_{pose} [-]	0.44 ± 0.06	0.35 ± 0.05	0.41 ± 0.06	0.46 ± 0.05	0.45 ± 0.06
	90°	E_t [m]	5.46 ± 6.34	0.92 ± 1.05	1.93 ± 2.00	0.93 ± 1.15	1.04 ± 1.28
		E_R [°]	23.62 ± 12.33	18.50 ± 9.29	24.52 ± 11.99	24.30 ± 13.00	28.73 ± 14.02
		E_{pose} [-]	0.48 ± 0.23	0.39 ± 0.21	0.39 ± 0.20	0.63 ± 0.33	0.64 ± 0.34



(a) SHIRT (TRON) (b) ASTRA (FRESCO) (c) ASTRA (COSMOS)

Figure 7: Domain interpolation between synthetic and target. Pose error is reported as α interpolates between the synthetic ($\alpha = 0$) and target ($\alpha = 1$) domains for (a) SHIRT with TRON, (b) ASTRA with FRESCO, and (c) ASTRA with COSMOS.

qualitatively in line with SHIRT real imagery while preserving pose-critical geometry such as edges and keypoint layout. This cross-dataset transfer suggests that FRESCO carries SHIRT real appearance to HST without retraining, motivating the domain-interpolation stress test that follows, while the COSMOS outputs deferred to the appendix exhibit substantially greater background variability and lens-effect diversity while remaining geometrically faithful to the edge conditioning.

Domain-interpolation stress test. To probe how pose accuracy changes along a controlled path from synthetic to real, we interpolate each frame as

$$x(\alpha) = (1 - \alpha)\, x_{\text{syn}} + \alpha\, x_{\text{real}}, \qquad \alpha \in [0, 1], \tag{4}$$

where x_{syn} is the synthetic image and x_{real} is the corresponding real image taken as the TRON capture for SHIRT and as either the FRESCO or COSMOS output for ASTRA. We identify the 100 most challenging frames per satellite at the real endpoint $\alpha = 1$ and trace their pose error as α increases from 0 to 1, with Figure 7 reporting the mean and standard deviation of both rotation error in degrees on the left axis and translation error in meters on the right axis. Under FRESCO the errors remain near their synthetic baseline up to $\alpha \approx 0.3$ and then rise sharply, exhibiting a clear loss barrier Lucas et al. (2021) at which small additional domain shifts produce disproportionately large error increases. Under COSMOS, by contrast, both errors grow already at small α and follow a smoother monotonic trajectory, consistent with the broader scene and illumination changes that the world model introduces in addition to the spectral shifts isolated by FRESCO. Together these curves indicate that FRESCO and COSMOS reproduce complementary failure signatures of real data and can serve as practical proxies for robustness diagnosis when paired real imagery is limited.

Method	FID ↓	CMMD ↓	Coverage ↑	Freq-L1 ↓	Edge-L1 ↓
UE5 raw	262.9	1.06	0.622	64.28	0.0223
COSMOS	**254.9**	**0.79**	**0.811**	54.11	0.0128
FRESCO	317.6	1.57	0.676	**7.81**	**0.0052**

Table 4: Distribution-distance comparison against 37 real HST photographs. Best per column in bold.

5.5 Distributional Comparison with Real Hubble Imagery

To assess how closely FRESCO and COSMOS approach real in-orbit imagery, we compare their outputs against 37 NASA photographs of the Hubble Space Telescope captured during servicing missions. We render 200 UE5 frames of HST, pass them through each pipeline to obtain matched COSMOS and FRESCO subsets, and convert all images to single-channel intensity to remove color bias. We report five metrics: FID on Inception-v3 features and CMMD Jayasumana et al. (2024) on CLIP ViT-B/32 features for high-level semantic alignment, PRDC Coverage Naeem et al. (2020) for the fraction of the real distribution recovered under nearest-neighbor matching, and the L1 distances between radially averaged Fourier magnitudes (Freq-L1) and between Sobel-edge histograms (Edge-L1) for spectral and structural agreement. Lower is better for all metrics except Coverage (Table 4).

COSMOS attains the lowest FID and CMMD and the highest Coverage, overlapping substantially with real HST imagery in feature space thanks to the diverse backgrounds, lens flares, and illumination styles introduced under edge conditioning. FRESCO, which modifies only the Fourier amplitude spectrum, instead attains order-of-magnitude lower Freq-L1 and substantially lower Edge-L1 than COSMOS, indicating tighter spectral and structural alignment with real photographs. That FRESCO does not improve over the raw rendering on FID and CMMD is consistent with its design, since these metrics aggregate semantic features outside its frequency-band parameterization. The two pipelines therefore offer complementary real-style coverage, justifying the release of both so that downstream methods can exploit either depending on the failure mode they target.

5.6 Onboard Inference on NVIDIA Jetson Platforms

Spacecraft pose estimation must run on board the chaser, since downlinking imagery is precluded during proximity operations, so deployment depends on sustaining real-time inference within tight latency, power, and thermal envelopes. We benchmark all five baselines on two NVIDIA Jetson generations that bracket current and next-generation small-satellite compute envelopes, namely the AGX Orin Developer Kit in the 30 W power mode and the Thor in MAXN mode. As reported in Table 5, Thor delivers 4–5× higher throughput than Orin across all five models, with YOLO-26n reaching 10.7 FPS on Thor against 2.5 FPS on Orin while SPNv2 is the slowest in both regimes due to its multi-task heads. On the 30 W Orin, only YOLO-26n approaches closed-loop frame rates, whereas Thor enables all five baselines to operate at usable rates. The energy per inference, in contrast, is comparable across the two platforms, indicating that the additional compute available on Thor is well utilized rather than wasted on idle overhead. The full setup, per-rail power and energy measurements, and thermal headroom analysis are deferred to Appendix I.

Table 5: Throughput (FPS) and energy per inference (J) on AGX Orin and Thor.

	FPS		Energy (J)	
Model	Orin	Thor	Orin	Thor
SPNv2	1.88	7.86	8.58	7.78
SPNv3	2.46	9.52	6.53	6.44
KRN	2.48	9.50	6.52	6.39
YOLO-26n	2.50	10.69	6.14	5.57
YOLO-26s	2.38	10.27	6.58	5.81

6 Conclusion

We introduced ASTRA, a large-scale multi-satellite benchmark for 6-DoF spacecraft pose estimation that covers twenty NASA-referenced targets under a unified UE5 rendering pipeline. Every target is distributed with synthetic sequential trajectories together with two complementary real-style variants generated by FRESCO and NVIDIA COSMOS, removing the dependence on physical HIL facilities while exposing the sim-to-real gap in a controllable way. Together, these components give the field a single benchmark with the target diversity, photometric realism, and uniform evaluation protocol it currently lacks. We see ASTRA as a foundation rather than an endpoint, with additional spacecraft and sim-to-real generators readily ingestible through the same pipeline.

AI USE STATEMENT

Generative AI tools were used to assist revising the manuscript. Specifically, they were used to improve the clarity, structure, and presentation of the text during the writing and editing process. The authors reviewed and verified all AI-assisted content and take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

We have made every effort to ensure that ASTRA and all reported results are reproducible. All twenty spacecraft are built from publicly available NASA 3D Resources CAD models (Appendix B). The full UE5 generation pipeline, covering the rendering setup, the ROE trajectory scenarios, and the per-frame annotation schema, is described in Appendix C. The FRESCO formulation is given in Section 4.1, with background in Appendices D and E. The COSMOS pipeline uses the publicly released COSMOS-Transfer2.5 model without fine-tuning, and all ten text prompts are listed in Table 8. Training hyperparameters, preprocessing, and augmentations are reported in Appendix F. The onboard benchmarking protocol, including power modes, software versions, and warm-up and measurement iterations, is detailed in Appendix I. The 37 real HST photographs used in Section 5.5 are public-domain NASA imagery (Appendix H). Upon publication, we will release the full dataset, the rendering and real-style generation code, the trained FRESCO weights, and the baseline training and evaluation scripts.

REFERENCES

Niket Agarwal, Arslan Ali, Maciej Bala, Yogesh Balaji, Erik Barker, Tiffany Cai, Prithvijit Chattopadhyay, Yongxin Chen, Yin Cui, Yifan Ding, et al. Cosmos world foundation model platform for physical ai. *arXiv preprint arXiv:2501.03575*, 2025.

Hansheng Chen, Pichao Wang, Fan Wang, Wei Tian, Lu Xiong, and Hao Li. Epro-pnp: Generalized end-to-end probabilistic perspective-n-points for monocular object pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2781–2790, 2022.

Woojin Cho, Junghwan Park, Sangcheol Sim, Steve Andreas Immanuel, Junhyuk Heo, and Darongsae Kwon. Elmzip: Onboard satellite image compression via extreme learning machines for efficient downlink. *arXiv preprint arXiv:2608.06942*, 2026.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pp. 1180–1189. PMLR, 2015.

Eberhard Gill, Simone D'Amico, and Oliver Montenbruck. Autonomous formation flying for the prisma mission. *Journal of Spacecraft and Rockets*, 44(3):671–681, 2007.

Gianluca Giuffrida, Lorenzo Diana, Francesco De Gioia, Gionata Benelli, Gabriele Meoni, Massimiliano Donati, and Luca Fanucci. Cloudscout: A deep neural network for on-board cloud detection on hyperspectral images. *Remote Sensing*, 12(14):2205, 2020.

Giorgia Guerrisi, Fabio Del Frate, and Giovanni Schiavon. Artificial intelligence based on-board image compression for the ϕ-sat-2 mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:8063–8075, 2023.

Jonáš Herec, Vít Růžička, and Rado Pitoňák. Optimizing methane detection on board satellites: Speed, accuracy, and low-power solutions for resource-constrained hardware. In *2025 European Data Handling & Data Processing Conference (EDHPC)*, pp. 1–9. IEEE, 2025.

Jiaxing Huang, Dayan Guan, Aoran Xiao, and Shijian Lu. Fsdr: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6891–6902, 2021.

Sadeep Jayasumana, Srikumar Ramalingam, Andreas Veit, Daniel Glasner, Ayan Chakrabarti, and Sanjiv Kumar. Rethinking fid: Towards a better evaluation metric for image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9307–9315, 2024.

George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.

Leela S Karumbunathan. Nvidia jetson agx orin series. *A Giant leap forward for robotics and edge AI applications. Technical Brief*, 2022.

Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pp. 2938–2946, 2015.

Mate Kisantal, Sumant Sharma, Tae Ha Park, Dario Izzo, Marcus Märtens, and Simone D'Amico. Satellite pose estimation challenge: Dataset, competition design, and results. *IEEE Transactions on Aerospace and Electronic Systems*, 56(5):4083–4098, 2020.

Adam W Koenig, Tommaso Guffanti, and Simone D'Amico. New state transition matrices for spacecraft relative motion in perturbed orbits. *Journal of Guidance, Control, and Dynamics*, 40 (7):1749–1768, 2017.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Rahul G Krishnan, Uri Shalit, and David Sontag. Deep kalman filters. *arXiv preprint arXiv:1511.05121*, 2015.

Taeyeop Lee, Jonathan Tremblay, Valts Blukis, Bowen Wen, Byeong-Uk Lee, Inkyu Shin, Stan Birchfield, In So Kweon, and Kuk-Jin Yoon. Tta-cope: Test-time adaptation for category-level object pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21285–21295, 2023.

Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Ep n p: An accurate o (n) solution to the p n p problem. *International journal of computer vision*, 81(2):155–166, 2009.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

James Lucas, Juhan Bae, Michael R Zhang, Stanislav Fort, Richard Zemel, and Roger Grosse. Analyzing monotonic linear interpolation in neural network loss landscapes. *arXiv preprint arXiv:2104.11044*, 1, 2021.

Eric Marchand, Hideaki Uchiyama, and Fabien Spindler. Pose estimation for augmented reality: a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12):2633–2651, 2015.

Sparsh Mittal. A survey on optimized implementation of deep learning models on the nvidia jetson platform. *Journal of Systems Architecture*, 97:428–442, 2019.

Mohamed Adel Musallam, Kassem Al Ismaeil, Oyebade Oyedotun, Marcos Damian Perez, Michel Poucet, and Djamila Aouada. Spark: Spacecraft recognition leveraging knowledge of space environment. 2021. doi: 10.48550/ARXIV.2104.05978. URL https://arxiv.org/abs/2104.05978.

Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh, Yunjey Choi, and Jaejun Yoo. Reliable fidelity and diversity metrics for generative models. In *International conference on machine learning*, pp. 7176–7185. PMLR, 2020.

Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 2005.

Junghwan Park, Sangcheol Sim, Woojin Cho, and Darongsae Kwon. Summarize first, download later: Onboard vlms for bandwidth-efficient earth observation. *arXiv preprint arXiv:2608.06959*, 2026.

Tae Ha Park and Simone D'Amico. Bridging domain gap for flight-ready spaceborne vision. *arXiv preprint arXiv:2409.11661*, 2024.

Tae Ha Park and Simone D'Amico. Adaptive neural-network-based unscented kalman filter for robust pose tracking of noncooperative spacecraft. *Journal of Guidance, Control, and Dynamics*, 46(9):1671–1688, 2023.

Tae Ha Park and Simone D'Amico. Online supervised training of spaceborne vision during proximity operations using adaptive kalman filtering. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11744–11752. IEEE, 2024a.

Tae Ha Park and Simone D'Amico. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. *Advances in Space Research*, 73(11):5726–5740, 2024b.

Tae Ha Park, Sumant Sharma, and Simone D'Amico. Towards robust learning-based pose estimation of noncooperative spacecraft. *arXiv preprint arXiv:1909.00392*, 2019.

Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In *2022 IEEE aerospace conference (AERO)*, pp. 1–15. IEEE, 2022.

Tae Ha Park, Marcus Märtens, Mohsi Jawaid, Zi Wang, Bo Chen, Tat-Jun Chin, Dario Izzo, and Simone D'Amico. Satellite pose estimation competition 2021: Results and analyses. *Acta Astronautica*, 204:640–665, 2023.

TH Park, J Bosse, and S D'Amico. Robotic testbed for rendezvous and optical navigation: Multi-source calibration and machine learning use cases, 2021 aas. In *AIAA Astrodynamics Specialist Conference, Big Sky*, 2021.

Nicola A Piga, Yuriy Onyshchuk, Giulia Pasquale, Ugo Pattacini, and Lorenzo Natale. Roft: Real-time optical flow-aided 6d object pose and velocity tracking. *IEEE Robotics and Automation Letters*, 7(1):159–166, 2021.

Vít Růžička, Gonzalo Mateo-Garcia, Luis Gómez-Chova, Anna Vaughan, Luis Guanter, and Andrew Markham. Semantic segmentation of methane plumes with hyperspectral machine learning models. *Scientific Reports*, 13(1):19999, 2023.

Ranjan Sapkota, Rahul Harsha Cheppally, Ajay Sharda, and Manoj Karkee. Yolo26: key architectural enhancements and performance benchmarking for real-time object detection. *arXiv preprint arXiv:2509.25164*, 2025.

Long Tian, Changjae Oh, and Andrea Cavallaro. Test-time adaptation for 6d pose tracking. *Pattern Recognition*, 152:110390, 2024.

Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 23–30. IEEE, 2017.

Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7167–7176, 2017.

Szabolcs Velkei, Csaba Goldschmidt, and Károly Vass. A large-scale, physically-based synthetic dataset for satellite pose estimation. *arXiv preprint arXiv:2506.12782*, 2025.

Qinwei Xu, Ruipeng Zhang, Ya Zhang, Yanfeng Wang, and Qi Tian. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14383–14392, 2021.

Yan Xu, Kwan-Yee Lin, Guofeng Zhang, Xiaogang Wang, and Hongsheng Li. Rnnpose: Recurrent 6-dof object pose refinement with robust correspondence field estimation and pose optimization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14880–14890, 2022.

Yanchao Yang and Stefano Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4085–4095, 2020.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3836–3847, 2023.

A Limitations and Broader Impact

ASTRA shares the principal limitation of every sim-to-real benchmark that operates without paired on-orbit imagery, namely that no synthetic or generated dataset can fully substitute for genuine in-orbit photographs, and the domain-interpolation results should be read as robustness diagnostics rather than as guarantees of operational performance. We close this gap as far as feasible through high-fidelity UE5 rendering and two complementary real-style pipelines that approach the photometric and structural statistics of real imagery from different directions. The twenty targets are also drawn from publicly redistributable NASA assets, which excludes classified or international platforms whose geometric and material properties may differ from the publicly released CAD models.

Releasing a multi-spacecraft pose-estimation benchmark also carries dual-use considerations, since the same models that enable on-orbit servicing and active debris removal can in principle support adversarial proximity operations against uncooperative spacecraft. We mitigate this concern by restricting ASTRA to publicly released NASA reference assets, while emphasizing that the dominant near-term applications of the work are sustainability-oriented, including life extension of existing assets, deorbit assistance for defunct platforms, and autonomous assembly of next-generation observatories. Future work will extend the target set, incorporate genuine on-orbit imagery as it becomes available, and embed the trained pose networks within closed-loop navigation pipelines.

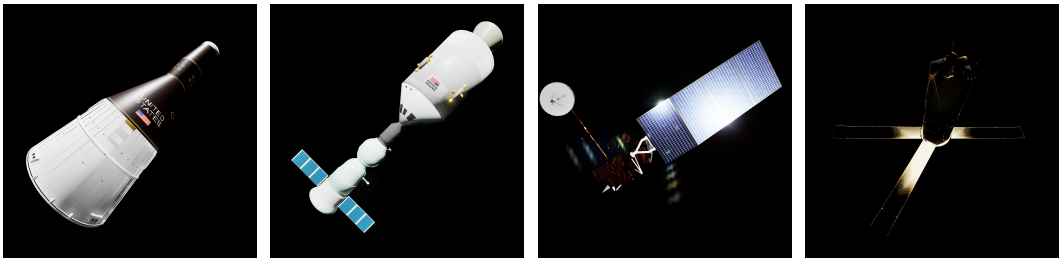
B Satellite List and Per-Target Statistics

This appendix provides the full list of the twenty spacecraft included in ASTRA together with their key attributes. Table 6 summarizes each target's mission category, operating agency, orbital regime, launch year, and the number of 3D keypoints N_{kpts} defined on its CAD model, while Figure 8 provides thumbnail renderings of all twenty targets with subfigure indices matching the table IDs.

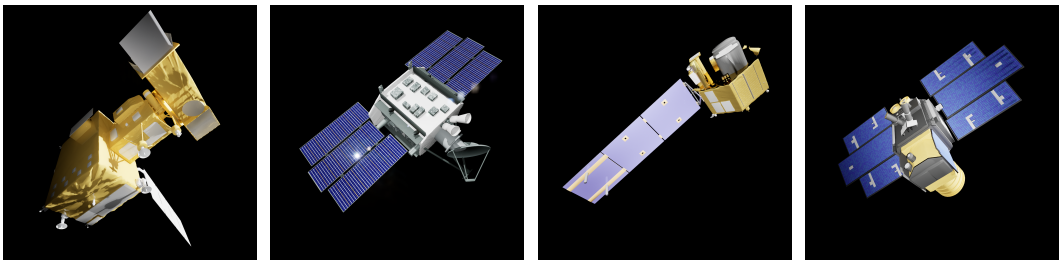
Table 6: Overview of the proposed satellite 3D model dataset. All 20 spacecraft models are obtained from the publicly available NASA 3D Resources repository. Entries are grouped by mission category and ordered chronologically by launch year within each group. N_{kpts} denotes the number of 3D keypoints defined on each target's CAD model.

ID	Spacecraft	Category	Agency	Domain	Year	N_{kpts}	Primary Mission
1	Gemini	Crewed	NASA	LEO	1965	13	Rendezvous & EVA
2	Apollo–Soyuz	Crewed	NASA/USSR	LEO	1975	30	Int'l crewed docking
3	Landsat 4/5	Earth Obs.	NASA/USGS	SSO LEO	1982	39	Land imaging
4	SeaStar	Earth Obs.	NASA	SSO LEO	1997	10	Ocean color
5	Landsat 7	Earth Obs.	NASA/USGS	SSO LEO	1999	39	Land imaging
6	QuikSCAT	Earth Obs.	NASA	SSO LEO	1999	27	Ocean winds
7	EO-1	Earth Obs.	NASA	SSO LEO	2000	22	Hyperspectral imaging
8	ICESat	Earth Obs.	NASA	Polar LEO	2003	31	Ice & elevation
9	OSTM/Jason-2	Earth Obs.	NASA/CNES	LEO	2008	26	Ocean topography
10	Landsat 8	Earth Obs.	NASA/USGS	SSO LEO	2013	30	Land imaging
11	GeoTail	Helio.	JAXA/NASA	Magnetotail	1992	16	Magnetosphere study
12	Parker Solar Probe	Helio.	NASA	Heliocentric	2018	27	Solar corona study
13	Voyager	Planetary	NASA	Interstellar	1977	20	Outer-planet flybys
14	Mars Odyssey	Planetary	NASA	Mars orbit	2001	24	Mars surface mapping
15	Rosetta	Planetary	ESA	Comet orbit	2004	28	Comet rendezvous
16	MRO	Planetary	NASA	Mars orbit	2005	22	Mars reconnaissance
17	Juno	Planetary	NASA	Jupiter orbit	2011	19	Jupiter polar study
18	Hubble	Telescope	NASA/ESA	LEO	1990	36	UV–NIR astronomy
19	JWST	Telescope	NASA/ESA/CSA	L2	2021	36	Infrared astronomy
20	Roman	Telescope	NASA	L2	2027*	18	Wide-field IR survey

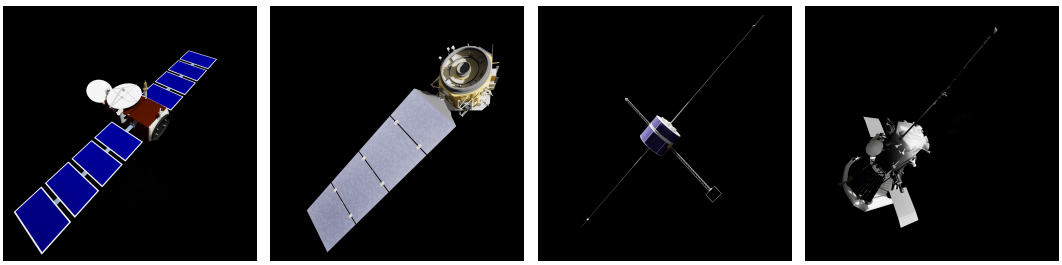
*Planned launch. Helio.: Heliophysics; SSO: Sun-synchronous orbit; LEO: Low Earth Orbit; L2: Sun–Earth second Lagrange point.



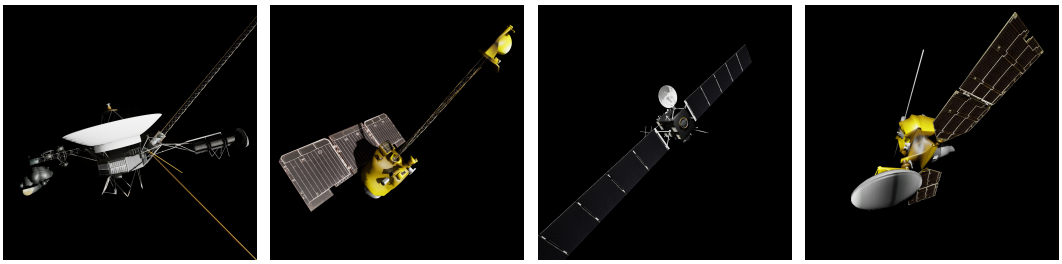
(a) Gemini (b) Apollo–Soyuz (c) Landsat 4/5 (d) SeaStar



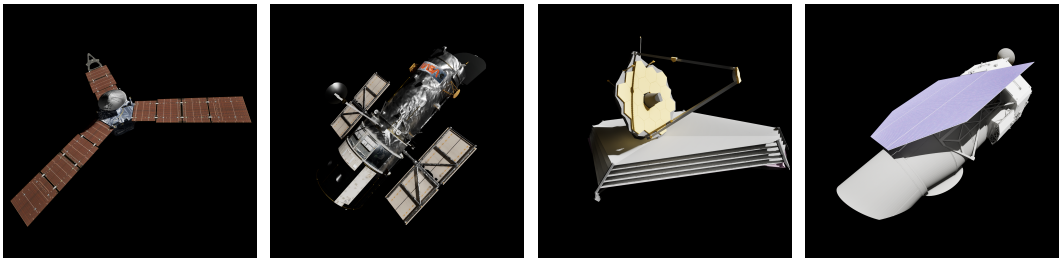
(e) Landsat 7 (f) QuikSCAT (g) EO-1 (h) ICESat



(i) OSTM/Jason-2 (j) Landsat 8 (k) GeoTail (l) Parker Solar Probe



(m) Voyager (n) Mars Odyssey (o) Rosetta (p) MRO



(q) Juno (r) Hubble (s) JWST (t) Roman

Figure 8: Visual overview of the 20 spacecraft 3D models in the proposed dataset. The indices (a)–(t) correspond to the IDs in Table 6.

C Detailed Dataset Generation Pipeline

This appendix describes how each ASTRA frame is generated end-to-end. The pipeline takes a NASA CAD asset, places it inside a virtual orbital scene with a chaser camera, and renders annotated frames through a single rendering backend shared across both data splits. We describe each stage in turn, starting with the rendering setup, then the trajectory model used for the test split, and finally the per-frame annotation procedure. The same pipeline is applied uniformly to all twenty targets.

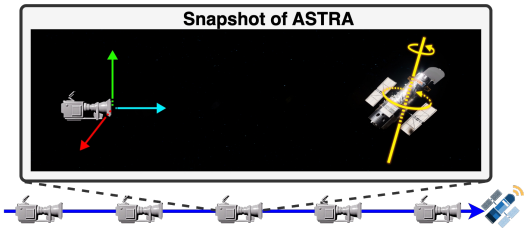
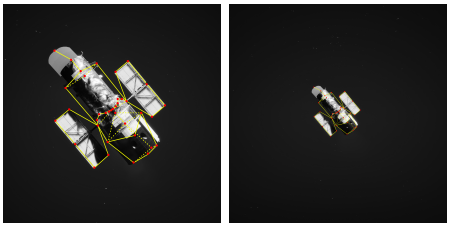


Figure 9: Overview of the ASTRA generation pipeline, which renders frames in UE5.

ASTRA separates training and test data into two distinct regimes. The training split consists of independent instance-level frames in which chaser pose, lighting direction, and Earth or Sun backgrounds are sampled independently per frame to maximize appearance diversity. The test split is generated sequentially along physically motivated rendezvous trajectories with no background, providing a clean evaluation regime in which the target geometry is isolated from photometric distractors and consecutive frames carry the temporal coherence characteristic of real proximity operations. The trajectories follow two Relative Orbital Elements (ROE) scenarios, namely ROE1 corresponding to a v-bar station-keeping hold and ROE2 corresponding to a terminal approach, which cover the two regimes that dominate practice in proximity operations. Every frame in both splits is annotated with the full 6-DoF pose and a per-target set of 3D keypoints whose counts are summarized in Table 6.

C.1 Rendering Setup

Each trajectory is rendered inside UE5 with the target imported from its NASA CAD asset and articulated where the source provides movable sub-components such as solar arrays and antennae. The scene contains a single directional sun light whose direction is sampled per trajectory and an Earth-albedo source whose intensity and direction follow the chaser's orbital position. Frames are captured by a virtual pinhole camera with two intrinsics, namely a narrow 45° field of view used for medium- to close-range inspection and a wide 90° field of view used for close-range inspection, with representative views of the HST subset under both settings shown in Figure 10. We additionally enable in-engine post-processing for lens flare, blooming, vignetting, and motion blur to approximate the photometric signature of real onboard sensors. All frames are output at 2048×2048 resolution and 24 frames per second.



(a) FOV 45° (b) FOV 90°

Figure 10: Example HST views under the two ASTRA camera fields of view.

C.2 Relative Motion and Trajectory Scenarios

ASTRA defines two trajectory scenarios that span the operationally relevant motion regimes for proximity operations. The naming convention follows SHIRT Park & D'Amico (2023), but our trajectories are independently generated and do not reuse SHIRT's parameterization. The overall flow of the simulator is summarized in Figure 9.

ROE1: V-bar hold. The chaser maintains a near-fixed relative position along the velocity vector of the target, often referred to as the v-bar direction. The relative position oscillates with small bounded amplitude around a nominal stand-off distance, and the viewpoint changes only slowly as the target rotates. ROE1 produces sequences with high temporal coherence, narrow viewpoint variation, and limited illumination change within a trajectory, which is representative of station-keeping and inspection phases. From a benchmark perspective, ROE1 is a controlled setting in which sequential models and filters can be evaluated under low motion diversity.

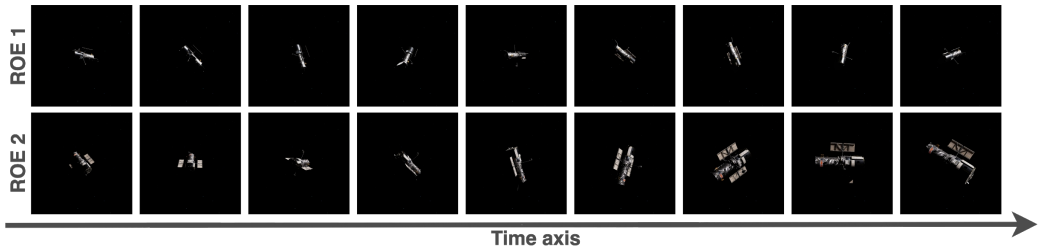


Figure 11: Representative frames from the two ASTRA trajectory scenarios, illustrated on the Hubble Space Telescope. ROE1 (v-bar hold) exhibits near-stationary relative motion with constrained viewpoints, while ROE2 (approach) exhibits broader variation in distance, viewpoint, and illumination.

ROE2: Approach. The chaser executes a continuous closing maneuver toward the target, with the relative range decreasing monotonically over the trajectory while the relative bearing varies smoothly. ROE2 sequences therefore exhibit large changes in apparent target scale, viewpoint, and illumination as the chaser transitions from far-field to near-field. This regime is representative of the terminal-approach phase of rendezvous and docking, and is the more challenging of the two for single-frame estimators because near-field frames produce strong perspective effects, occlusions of articulated sub-components, and saturation under direct sunlight.

For each target and each scenario we render 10 trajectories of 240 frames at both fields of view, sampling the chaser's initial position, velocity, and rotational state, the target's rotation rate and axis, and the global Sun direction independently per trajectory so that no two trajectories share the same illumination and motion configuration. Figure 11 shows representative frames from both scenarios on the Hubble Space Telescope.

C.3 Per-Frame Annotation

A target-specific set of up to 39 keypoints is defined once per target on its CAD model in the target's body frame, with the exact number varying per spacecraft according to its structural complexity (see Table 6). At each frame the keypoints are projected to image coordinates using the simulator's known camera matrix and target pose, and visibility is determined from the UE5 depth buffer so that occluded or off-screen keypoints are explicitly flagged. Each frame is then written out together with a JSON file containing the 6-DoF pose as a quaternion and translation, the visible keypoint coordinates, the camera intrinsics, the sun-vector and Earth-albedo state, and a record of the rendering parameters used. The schema is identical across all twenty targets so that any target's annotations can be consumed without format changes.

D Fourier Domain Adaptation: Background and Connection to FRESCO

This appendix reviews Fourier Domain Adaptation (FDA) Yang & Soatto (2020) and explains how it motivates the design of FRESCO (Section 4.1). The Fourier Neural Operator framework that underpins FRESCO's learnable filter is reviewed separately in Appendix E.

Amplitude–phase disentanglement. For an image I defined on $\Omega \subseteq \mathbb{R}^2$, the 2D Fourier transform $\hat{I}(k) = \mathcal{F}[I](k)$ decomposes uniquely as $\hat{I}(k) = |\hat{I}(k)| \cdot e^{j\angle \hat{I}(k)}$, and the two components carry qualitatively different information Oppenheim & Lim (2005). Phase encodes the relative spatial alignment of frequency components and acts as the primary carrier of geometric structure such as edges, corners, and contours. Amplitude encodes the energy distribution across frequencies and correlates with global photometric style, with low-frequency amplitude tracking illumination and color balance and high-frequency amplitude tracking texture and sensor noise. A transformation that modifies amplitude but leaves phase intact therefore changes appearance without disturbing the geometric cues that pose estimators rely on.

The FDA operator. Given a synthetic image I_s and a real image I_r, FDA replaces the low-frequency amplitude of the source with that of the target while keeping the source phase. Let $M_\beta(k) \in \{0,1\}$ be a binary mask that is 1 on a centered low-frequency disc of relative radius $\beta \in [0,1]$ around the DC component and 0 elsewhere. The FDA-transformed image is then given by

$$I_{s\to r} = \mathcal{F}^{-1}\Big[\big(M_\beta \odot |\hat{I}_r| + (1 - M_\beta) \odot |\hat{I}_s|\big) \cdot e^{j\angle \hat{I}_s}\Big]. \tag{5}$$

Because the phase is preserved by construction, $I_{s\to r}$ retains the geometry of I_s while inheriting the global appearance of I_r inside the band. The single hyperparameter β controls how aggressive the style transfer is. Small β transfers only illumination and color balance, while larger β also rewrites mid-frequency amplitudes at the cost of artifacts near the band boundary, since the binary mask M_β introduces a sharp discontinuity in the spectrum.

Limitations in the spacecraft setting. FDA is attractive because it is training-free and preserves geometry by construction, but it has concrete limitations for spacecraft pose estimation. The mask M_β is binary and isotropic, treating all directions and all frequencies inside the band identically, whereas the gap between synthetic and real spacecraft imagery is anisotropic (e.g., specular highlights on solar panels) and frequency-dependent (illumination at low frequencies versus sensor noise at high frequencies). The target amplitude $|\hat{I}_r|$ is also sampled from a single randomly chosen real image, so the transferred style is noisy and not aligned with any learned notion of the real distribution.

From FDA to FRESCO. FDA's amplitude replacement in Eq. equation 5 is equivalent to applying a per-mode multiplicative filter to the source spectrum, expressed as

$$\hat{I}_{s\to r}(k) = R_r^{\text{FDA}}(k;\beta) \cdot \hat{I}_s(k), \qquad R_r^{\text{FDA}}(k;\beta) = \begin{cases} |\hat{I}_r(k)|/|\hat{I}_s(k)| & k \in \mathcal{B}_\beta, \\ 1 & \text{otherwise.} \end{cases} \tag{6}$$

The filter values are determined by a single random real image and are supported on a hard low-frequency disc $\mathcal{B}_\beta$. FRESCO replaces this filter with a learnable, band-restricted Fourier filter $R_\varphi(k;\rho)$ trained on a small paired set of synthetic and real images. The resulting operator recovers FDA as a degenerate special case while removing the sharp band boundary, supporting frequency-dependent edits, and aggregating style information from many synthetic and real pairs. The full formulation is given in Section 4.1, and the underlying neural operator framework is summarized next in Appendix E.

E Neural Operator and FNO Background

This appendix gives the formal background on Neural Operators (NO) and the Fourier Neural Operator (FNO) Li et al. (2020) that underpin the learnable spectral filter $R_\varphi(k;\rho)$ used in FRESCO (Section 4.1). Whereas Appendix D writes FDA as a fixed, image-specific multiplicative filter in the frequency domain, FNO provides a principled construction in which such a filter becomes the learnable component of a neural network layer.

Neural operators. Let $\mathcal{A}$ and $\mathcal{U}$ be separable Banach spaces of functions defined on a domain $\Omega \subseteq \mathbb{R}^2$. A neural operator Lu et al. (2021); Li et al. (2020); Karniadakis et al. (2021); Kovachki et al. (2023) approximates a target operator $\mathcal{G}^\dagger : \mathcal{A} \to \mathcal{U}$ through a parameterized integral

$$(\mathcal{G}a)(x) = \int_\Omega \kappa(x,y)\, a(y)\, dy, \tag{7}$$

where $\kappa(x,y)$ is a learnable kernel. Stacked operator layers are realized through the iterative update

$$v_{i+1}(x) = \sigma(W\, v_i(x) + (\mathcal{K}_\varphi v_i)(x))\,, \tag{8}$$

where W is a pointwise linear transform, σ a non-linear activation, $\mathcal{K}_\varphi$ the kernel integral operator, and v_i the intermediate representation at iteration i.

Fourier Neural Operator. FNO Li et al. (2020) realizes the kernel integral operator efficiently in Fourier space as

$$(\mathcal{K}_\varphi v_i)(x) = \mathcal{F}^{-1}(R_\varphi(k) \cdot \mathcal{F}[v_i](k))(x), \tag{9}$$

where $R_\varphi(k)$ is a learnable Fourier filter and k indexes discrete frequency modes. Because pointwise multiplication in the frequency domain corresponds to convolution in the spatial domain, this parameterization captures long-range, global interactions through a small set of spectral parameters, and is well suited to learning transformations that act differently across frequency bands while leaving the geometric structure carried by the phase intact.

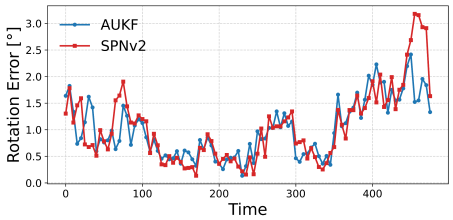
From FNO to FRESCO. FRESCO inherits the FNO parameterization but specializes it to the sim-to-real style-transfer setting. Two changes are introduced. First, the learnable filter $R_\varphi(k;\rho)$ is restricted to a low- and mid-frequency band of radius ρ around the spectral origin, since the gap between synthetic and real spacecraft imagery concentrates in this regime Oppenheim & Lim (2005). Second, the filter is applied only to the amplitude, with the phase of the synthetic input left unchanged. The resulting operator can be viewed as a learnable, smooth generalization of the FDA filter $R_r^{\text{FDA}}(k;\beta)$ in Eq. equation 6: the binary disc support $\mathcal{B}_\beta$ is replaced by a soft band of radius ρ, and the per-image amplitude ratios are replaced by parameters trained on many synthetic and real pairs. The full FRESCO formulation is given in Section 4.1.

F Training Details

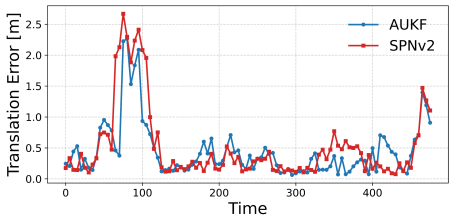
All models are trained with the AdamW optimizer using a learning rate of 5×10^{-4}, weight decay of 1×10^{-4}, and momentum coefficients $(\beta_1, \beta_2) = (0.9, 0.999)$. The learning rate is decayed by a factor of 0.15 at epochs 10, 15, and 18, and mixed-precision FP16 is used throughout. SPNv2, SPNv3, and KRN share the EfficientDet-D3 backbone with its accompanying BiFPN feature pyramid, while YOLO-26n and YOLO-26s use their native YOLO backbones initialized from COCO-pretrained weights. The output heads follow the original publications. Input frames are resized to 1024×1024 before being passed to the network. All images are converted to single-channel intensity, replicated across three channels to maintain compatibility with ImageNet-pretrained backbones, and normalized using the standard ImageNet mean and variance statistics. During training we apply standard photometric augmentations including brightness and contrast jitter, motion and median blur, additive Gaussian noise, sun-flare overlay, gamma correction, and HSV perturbation, sampled independently per frame to encourage robustness to the photometric variability observed in real onboard imagery.

G Sequential Pose Estimation: AUKF Filtering on ASTRA

A central design choice of ASTRA is that every target is distributed as sequential trajectories rather than independent frames, since real proximity-operations workloads are inherently sequential and many practical pose estimators couple a per-frame predictor with a temporal filter. To illustrate how sequential information can be exploited on ASTRA, we layer the Adaptive Unscented Kalman Filter (AUKF) Park & D'Amico (2024a) on top of SPNv2 Park & D'Amico (2024b) and evaluate on the FRESCO real-style test set, where the per-frame predictor faces an appearance shift relative to its synthetic training distribution. SPNv2 produces a per-frame pose estimate, and the AUKF maintains the chaser–target relative pose and angular velocity with adaptive process- and measurement-noise covariances driven by the SPNv2 residual statistics.



(a) Rotation error



(b) Translation error

Figure 12: Per-frame errors on an ASTRA FRESCO trajectory, with and without AUKF.

Figure 12 shows the per-frame rotation and translation errors along the trajectory. SPNv2 alone exhibits intermittent error spikes at frames with challenging viewpoints, occlusions, or specular highlights, while AUKF attenuates these spikes by

leveraging temporal coherence and the slow time scale of the chaser–target relative motion. In aggregate across the twenty targets the filter reduces error variance over each trajectory while leaving the trajectory-averaged mean error largely unchanged, indicating that AUKF reshapes the error distribution rather than the systematic bias of the per-frame predictor. The persistence of non-trivial residual error after filtering also suggests that ASTRA's sequential trajectories leave room for more capable temporal models such as RNN-based trackers Xu et al. (2022), flow-assisted filters Piga et al. (2021); Krishnan et al. (2015), and test-time adaptation methods Lee et al. (2023); Tian et al. (2024).

H Real Hubble Imagery Used for Distributional Validation

The distributional comparison reported in Section 5.5 uses a set of 37 publicly available NASA photographs of the Hubble Space Telescope downloaded from the NASA image library and captured during on-orbit servicing missions as well as during free-flying configurations photographed from accompanying Space Shuttle missions. The collection spans a wide range of viewing conditions, with the camera-to-target distance ranging from close-range inspection captured during astronaut spacewalks to far-range photographs acquired from the receding shuttle, and with the background composition ranging from a uniform dark sky to a fully illuminated Earth-limb. This heterogeneity is essential for the validation, since it forces the distributional metrics in Table 4 to aggregate over realistic operational variability rather than over a single canonical pose. Figure 13 shows four representative examples that illustrate this diversity.

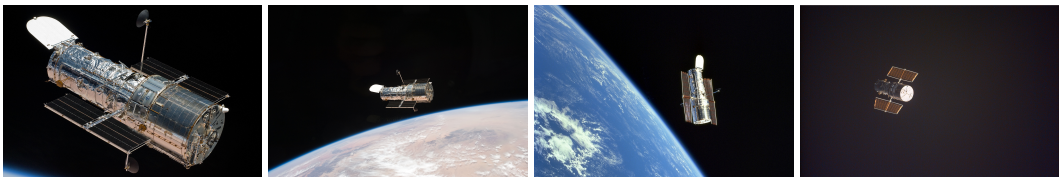


Figure 13: Four representative examples drawn from the 37-image validation set, illustrating the diversity in viewing distance, illumination, and background.

For consistency with the synthetic and generated subsets, all photographs were converted to three-channel grayscale by replicating the luminance channel across the red, green, and blue components so that residual color bias does not dominate the distributional comparison, and no further geometric or photometric processing was applied prior to metric computation. All photographs are public-domain NASA imagery sourced from `https://science.nasa.gov/mission/hubble/multimedia/hubble-images/` and reused under NASA media usage guidelines.

Table 7: Onboard inference benchmark on Jetson AGX Orin (30 W mode) and Jetson Thor (MAXN).

Platform	Model	Params (M)	Latency (ms)	FPS	Power (W)	Energy (J)	Peak T_j (°C)
AGX Orin	SPNv2	12.96	531.3	1.88	16.14	8.58	51.0
	SPNv3	12.26	406.6	2.46	16.06	6.53	52.0
	KRN	12.34	403.5	2.48	16.15	6.52	52.6
	YOLO-26n	5.07	400.0	2.50	15.34	6.14	52.7
	YOLO-26s	12.02	420.5	2.38	15.64	6.58	53.0
Thor	SPNv2	12.96	127.3	7.86	61.08	7.78	46.1
	SPNv3	12.26	105.1	9.52	61.33	6.44	46.7
	KRN	12.34	105.2	9.50	60.73	6.39	47.1
	YOLO-26n	5.07	93.5	10.69	59.60	5.57	46.8
	YOLO-26s	12.02	97.3	10.27	59.70	5.81	47.0

I ONBOARD INFERENCE ON NVIDIA JETSON PLATFORMS

Spacecraft pose estimation is ultimately performed on board the chaser, since downlinking imagery to ground is precluded by bandwidth, latency, and autonomy requirements during proximity operations. Practical deployment therefore depends not only on accuracy but also on whether each model can sustain real-time inference within the latency, power, and thermal envelopes available on flight-grade embedded hardware. To characterize this regime, we benchmark all five baselines on two NVIDIA Jetson generations that bracket the compute envelopes most representative of current and next-generation small-satellite payloads. The first platform is the Jetson AGX Orin Developer Kit operated in the 30 W power mode (`MODE_30W` in the `nvpmodel` configuration), which captures the constrained-power regime typical of current small-satellite missions, and the second is the Jetson Thor in MAXN mode, which represents the higher compute envelope of upcoming orbit-grade accelerators. All inference uses PyTorch 2.11 with the platform-native CUDA stack (12.6 on Orin and 13.0 on Thor), batch size one, and full FP32 precision without TensorRT optimization, on a 1024×1024 random input with 30 warm-up forwards followed by 200 measured iterations under explicit GPU synchronization at every step. Board power and junction temperature are sampled at 1 Hz from tegrastats throughout each run.

Table 7 reports the resulting latency, throughput, mean board power, energy per inference, and peak junction temperature for each model on both platforms. Thor delivers roughly four to five times higher throughput than Orin across all five models while remaining more than $25°$C below the SoC throttle threshold, whereas Orin operates at substantially lower mean power, around 16 W against around 60 W on Thor, but at a correspondingly lower frame rate of about 2.5 FPS against 10 FPS. The energy per inference, which combines latency and power into the most directly deployment-relevant metric, is comparable across the two platforms and is in fact slightly lower on Thor for the lightweight YOLO-26n variant, indicating that the additional compute on Thor is well utilized rather than wasted on idle overhead. Across architectures, YOLO-26n achieves the lowest latency and energy per inference on both platforms thanks to its smaller backbone, SPNv2 incurs the highest cost due to its multi-task heads, and SPNv3 together with KRN sits between the extremes with nearly identical compute footprints that reflect their shared EfficientDet-D3 backbone dominating the inference cost. Under the constrained 30 W envelope of Orin only YOLO-26n reaches the throughput required for closed-loop frame-rate operation, whereas the higher envelope of Thor enables all five baselines to operate at usable rates. We additionally note that all measurements use FP32 inference without TensorRT or quantization, so deployment with TensorRT FP16 or INT8 would reduce the latency and energy reported here substantially.

J COSMOS PIPELINE

This appendix describes the COSMOS-based real-style generation pipeline used in ASTRA. We first specify the model variant and the conditional generation procedure that maps UE5 edge sequences to photorealistic video, and then detail the text prompts used to control the photometric attributes of the generated frames.

J.1 CONDITIONAL GENERATION SETUP

We employ NVIDIA COSMOS-Transfer2.5[3], a video-to-video world model that generates photorealistic sequences conditioned on a structural input together with a free-form text prompt. The model is used out of the box without any fine-tuning on satellite imagery, which keeps the pipeline simple to reproduce. Figure 14 summarizes the pipeline as we use it for ASTRA.

For each test trajectory, the corresponding sequence of edge maps is paired with a text prompt that specifies the orbital context and the desired photometric attributes, and supplied to COSMOS-Transfer2.5 as the structural conditioning input. The model then synthesizes a photorealistic video sequence in which every frame inherits the projected geometry of the conditioning edge map at the same temporal index, so the per-frame 6-DoF pose and 3D keypoint annotations from the original UE5 render carry over without modification. Because the conditioning is supplied as a contiguous video rather than as independent frames, the generated sequence preserves temporal coherence

[3] `https://docs.nvidia.com/cosmos/latest/transfer2.5/index.html`

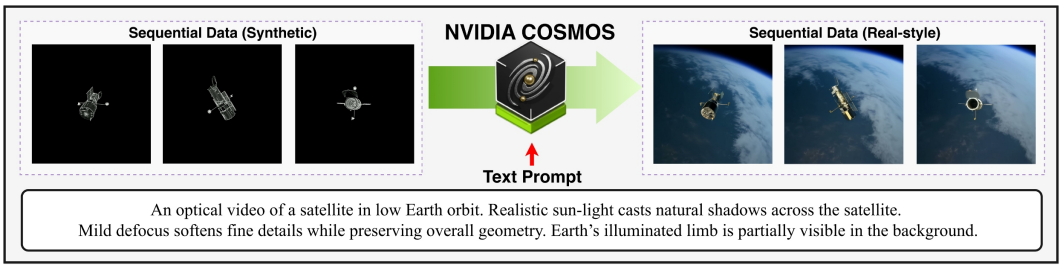


Figure 14: Overview of the COSMOS-Transfer2.5 pipeline used in ASTRA.

across consecutive frames, which is essential for evaluating sequential trackers on the ASTRA test split.

J.2 COSMOS Text Prompts

The COSMOS pipeline generates each photorealistic frame from a UE5-derived edge map together with a text prompt. To cover the range of orbital lighting and background conditions found in real on-orbit imagery, we use ten fixed prompt templates and sample one prompt per generated sequence. Each prompt varies along three axes. The first is the dominant illumination regime, ranging from direct sunlight to terminator twilight, backlight, gradient lighting, and minimal illumination. The second is the background content, which spans deep space with stars, the Earth limb, cloud-covered Earth, and silhouettes against the Sun. The third is the sensor or optical effects to be reproduced, such as specular reflections, sensor noise, motion blur, overexposed highlights, CCD bloom, and defocus. The full prompt set is listed in Table 8.

Table 8: Text prompts used to drive the COSMOS world model for ASTRA real-style generation. Each prompt is paired with a UE5-derived edge map at generation time.

ID	Photometric Regime	Prompt
0	Natural sunlight, star background	A high-fidelity optical video of a satellite floating in orbit. Realistic sunlight illumination casts natural shadows on the satellite. The background is the blackness of deep space with distant stars.
1	Strong direct light, specular reflections	A realistic optical observation of a satellite in space. Intense direct sunlight produces high-contrast lighting with bright specular reflections on the satellite surface. A clean star field stretches across the dark background.
2	Low light near terminator, sensor noise	An optical video of a satellite drifting near the orbital terminator. Dim illumination and reduced surface visibility create a moody, low-light scene. Sensor noise and grain are faintly present due to high-gain camera settings. Only scattered stars are visible in the surrounding darkness.
3	Earth-limb background, soft natural light	An optical video of a satellite orbiting above Earth. Sunlight illuminates the satellite with physically accurate reflections. Earth's bright limb curves along the lower portion of the frame, adding soft ambient light to the scene.
4	Backlight silhouette	An optical capture of a satellite against the blackness of space. The Sun is behind the satellite, creating a dramatic silhouette with bright rim lighting along its edges. Faint stars dot the dark background.
5	Uneven illumination, large Earth background	An optical video of a satellite in low Earth orbit. Uneven solar lighting leaves portions of the satellite brightly lit while other regions fall into deep shadow. The bright, cloud-covered Earth fills part of the background, creating strong contrast.
6	Overexposure with motion blur	A realistic optical observation of a satellite in orbit. Extremely strong solar illumination causes overexposed highlights on the satellite surface. Slight motion blur is visible due to the satellite's rotational movement. Deep space with distant stars fills the background.
7	Gradient illumination, cloud background	An optical video of a satellite in orbit above the Earth. Transitional lighting creates a gradient across the satellite, with one side in full sunlight and the other in shadow. Below, Earth's cloud-covered surface reflects sunlight into the scene.
8	Minimal illumination, CCD bloom	An optical capture of a satellite in the darkness of space. Minimal ambient light renders the satellite mostly dark, with only edges faintly visible. Bright specular points produce CCD bloom and faint charge smear. A sparse star field is visible in the deep black background.
9	Natural sunlight, mild defocus, Earth limb	An optical video of a satellite in low Earth orbit. Realistic sunlight casts natural shadows across the satellite. Mild defocus softens fine details while preserving overall geometry. Earth's illuminated limb is partially visible in the background.