Sequential Dataset for Satellite Pose Estimation and a Frequency-Space Neural Operator for HIL-Free Generalization Benchmarking

Woojin Cho[1*] Junghwan Park[1*] Steve Andreas Immanuel[1]
Junmin Park[1] Seokhyun Chin[2] Jiayun Wang[3]

[1]TelePIX [2]California Institute of Technology [3]Georgia Institute of Technology

Abstract

Accurate six-degree-of-freedom (6-DoF) pose estimation of non-cooperative satellites is critical for the success of on-orbit servicing, assembly, and manufacturing missions. However, the development of deep learning models is severely hampered by a significant scarcity of labeled, real-world on-orbit imagery, which creates a substantial sim-to-real domain gap. While existing datasets are valuable, they lack the large-scale, diverse sequential data necessary to develop and rigorously validate tracking, filtering, and test-time adaptation algorithms under realistic orbital dynamics. To address this sequential data gap, we introduce a new large-scale, public benchmark: ASTRA-HST, the Hubble Space Telescope (HST) sequential dataset. It consists of 512 unique and physically plausible rendezvous trajectory sequences of the geometrically complex HST. The dataset was generated via a high-fidelity simulator with a wide range of parameters, including orbital dynamics, illumination conditions, and camera properties, providing a new resource for the research community. To tackle the fundamental sim-to-real problem, we reframe domain adaptation as a function space mapping problem. We propose the frequency-space neural operator (FRESCO), a novel architecture that learns to translate synthetic images to the real domain by operating on distinct frequency bands of the Fourier amplitude spectrum while preserving the phase, which encodes geometric structure. We benchmark several state-of-the-art methods on ASTRA-HST and demonstrate how FRESCO, trained on existing hardware-in-the-loop (HIL) data, can be used to generate a realistic testbed for quantifying the sim-to-real gap on our new dataset.

1. INTRODUCTION

Robust 6-DoF pose estimation [1, 8] of a satellite is a crucial capability for autonomous spacecraft operations in orbit. Unlike terrestrial vision tasks, spaceborne pose estimation faces unique challenges.

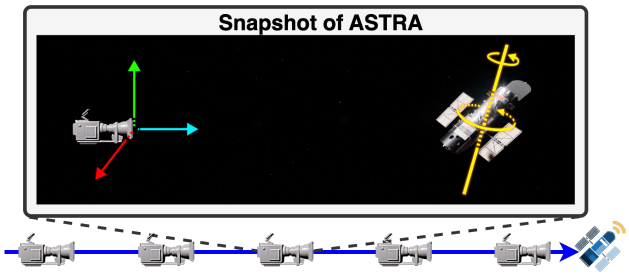


Figure 1. Generating ASTRA-HST dataset using a high-fidelity Unreal Engine simulator. The system varies orbital dynamics, camera parameters, Camera field of view (FOV), and output time-stamped frames with 6-DoF poses and keypoint annotations.

As acquiring large-scale labeled image datasets of a target spacecraft in its true space environment is impractical, researchers rely heavily on computer-based simulated images [27, 31]. While these synthetic images are easy to generate in bulk, they often fail to replicate the visual complexity and illumination conditions of real imagery. This discrepancy causes learned models to degrade in performance when deployed on actual images—a problem known as the sim-to-real domain gap [5, 33]. Prior efforts have attempted to bridge this gap by introducing hardware-in-the-loop (HIL) testbeds, such as the Testbed for Robotic Optical Navigation (TRON) facility [3, 21], which can capture images of a spacecraft mockup under space-like lighting for validation. However, data collection within HIL facilities is both limited in scale and expensive. A second challenge is that spacecraft rendezvous and docking are inherently sequential, temporal processes. A servicer spacecraft must continuously track a target's pose over time as it approaches or orbits, processing a stream of images in real time. Yet, most existing public datasets [9, 27] for satellite pose estimation consist of independent, static images without temporal continuity. Such issues are significant obstacles for re-

*Equal contribution

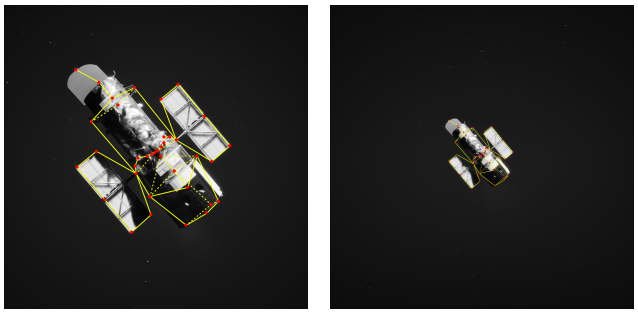
search as they force researchers to train models on limited, static single-image datasets and then resort to techniques like online learning or filtering at test time to adapt to temporal input. The gap hinders the development of algorithms that fully exploit dynamics and history in pose estimation.

In this paper, we address both the data scarcity and the sim-to-real gap with two contributions. First, as shown in Fig. 1, we present a new Hubble Space Telescope (HST) Sequential Pose Estimation Dataset, a collection of synthetic video sequences capturing diverse rendezvous trajectories with a detailed spacecraft model. To our knowledge, this is the largest and most varied public sequential dataset for satellite pose estimation. The HST sequences vary key parameters such as approach velocity, target spin rate, relative orientation axis, camera field-of-view, and initial distance to span a wide range of realistic scenarios. Each frame is annotated with the full 6-DoF pose and up to 36 keypoint projections, enabling both direct pose regression and keypoint-based pose algorithms. Second, we propose a FREquency-SpaCe neural Operator (FRESCO) for sim-to-real adaptation, a novel learning-based framework that maps synthetic images to a "realistic" domain in function space. FRESCO learns a complex frequency-domain filter to transform rendered images into a high-fidelity style that mimics real sensor images without altering the geometric content. By operating in the Fourier domain, our approach preserves phase information (which encodes pose-critical structure) while adjusting the amplitude spectrum in a frequency-dependent manner to inject realistic textures and noise. This enables us to generate real-style images from unlimited synthetic data, effectively substituting physical testbeds like TRON. Crucially, our neural operator (NO) learns this mapping across frequency bands, which allows it to capture global illumination shifts. We validate these contributions with extensive experiments. We train standard satellite pose estimation baselines on our synthetic HST sequences and evaluate them on FRESCO-generated images that simulate a realistic domain.

The results quantify the remaining domain gap and how our frequency-based adaptation improves robustness. We also conduct a gradual domain blending test by continuously mixing synthetic and generated real-style image attributes to analyze how performance degrades as the input distribution shifts. These experiments demonstrate that the proposed dataset and FRESCO can serve as scalable, low-cost tools to advance spacecraft pose estimation research.

The contributions of this paper are threefold:

- **ASTRA framework and dataset.** We introduce ASTRA, a general framework for generating sequential datasets from 3D spacecraft models. As a first instantiation, we release ASTRA-HST, comprising 512 unique and physically plausible rendezvous sequences of the geometrically complex HST, each with 960 frames, multi-FOV settings, and dense annotations.
- **FRESCO for real-style adaptation.** We present FRESCO, a frequency-space neural operator that translates synthetic images into realistic counterparts while preserving geometry and labels, enabling controlled sim-to-real analysis without expensive hardware-in-the-loop data collection.
- **HIL-free training/evaluation pipeline.** We propose a decoupled pipeline that leverages FRESCO to synthesize large-scale real-style training and test sets, supporting robust training and fair evaluation of downstream pose networks without requiring physical HIL setups.



(a) FOV 45° (b) FOV 90°

Figure 2. Example views under two camera fields-of-view. (a) Standard FOV 45° for medium- to close-range inspection; (b) Wide FOV 90° for close-range inspection.

2. Related Work

We briefly review prior work on spacecraft pose estimation datasets and sim-to-real domain adaptation, which together motivate the design of ASTRA-HST and FRESCO.

2.1. Spacecraft Pose Estimation Datasets

A central challenge in spacecraft pose estimation is the lack of labeled, in-orbit imagery, which has motivated the community to develop a series of synthetic and hardware-in-the-loop (HIL) datasets. The SPEED dataset [9] introduced the Tango spacecraft model from the PRISMA mission [3], combining $15k$ synthetic OpenGL renderings with 305 images captured in the TRON facility, thereby providing one of the public resources for benchmarking learning-based methods [25]. To extend realism, SPEED+ [27] added $60k$ synthetic images together with $9.5k$ real (HIL) data samples under two calibrated illumination domains—lightbox and sunlamp—captured at TRON, establishing a foundation for the international Satellite Pose Estimation Challenge (SPEC2021) [28].

The SPARK dataset [19] further broadened object diversity by providing synthetic renderings of 11 different spacecraft, supporting research on generalization across geometries [22]. More recently, the SHIRT dataset [23] intro-

Table 1. Comparison of Public Satellite Pose Estimation Datasets. ASTRA-HST provides the largest set of diverse sequential trajectories for a complex target (HST) with multi-FOV imagery and keypoints, enabling temporal benchmarking.

Dataset Name	Year	Target	Total Images	Test Data	Sequential	#Trajectories	#FOV	Keypoints	Resolution
SPEED	2019	Tango	∼15k	TRON	No	N/A	1	Yes	1920×1200
SPEED+	2021	Tango	∼70k	TRON	No	N/A	1	Yes	1920×1200
SHIRT	2022	Tango	∼5k	TRON	Yes	2	1	Yes	1920×1200
SPARK (stream 1)	2021	-	∼150k	N/A	No	N/A	1	No	1440×1080
SPARK (stream 2)	2021	Proba-2	∼150k	N/A	Yes	100	1	No	1440×1080
DLVS3	2025	HST	∼1,000k	N/A	No	N/A	1	No	1024×1024
ASTRA-HST	**2025**	**HST**	**∼1,000k**	**FRESCO**	**Yes**	**512**	**2**	**Yes**	**2048×2048**

duced sequential imagery for the first time, covering two short rendezvous scenarios (ROE1 with a *v-bar hold* trajectory and ROE2 with an *approaching* trajectory) with paired synthetic and TRON-captured sequences [24]. While this marked an important step toward temporal benchmarking, SHIRT remains limited by its small number of trajectories (only two), restricting its capacity to support large-scale sequential learning. The DLVS3 [34] is a fully-synthetic HST pose estimation dataset with 1000k images and customized materials using the MaterialX library.

Table 1 provides a detailed comparison of these datasets, highlighting differences in scale, realism, and temporal continuity. In contrast, our contribution provides a large-scale sequential dataset (ASTRA-HST) that focuses on a highly detailed spacecraft (HST) and covering diverse motion profiles, and introduces a neural operator that generates realistic images without requiring physical facilities.

We believe this dataset will enable new research into sequential pose estimation approaches, including RNN-based trackers [37], filtering strategies such as flow-assisted Kalman/UKF methods [12, 29], and test-time adaptation techniques applied to sequential pose data [13, 32].

2.2. Domain Adaptation for Spacecraft Pose Estimation

To bridge the sim-to-real gap, various domain adaptation techniques have been developed. While methods like domain randomization [33] and adversarial feature alignment have shown success, they often operate in the pixel space. An alternative and highly influential line of work operates in the frequency domain, with Fourier Domain Adaptation (FDA) [38] being a prime example. Understanding the motivation behind FDA is crucial to appreciating the design of our proposed method.

The core premise of using the Fourier domain is the disentanglement of properties. The 2D Fourier Transform decomposes an image into its constituent frequencies. Crucially, the amplitude spectrum and the phase spectrum of this transformation encode different types of information.

Phase Spectrum The phase contains the majority of the information about the spatial structure of the image—the location of edges, corners, and other critical geometric features. Preserving the phase is paramount to preserving the identity and pose of the object in the image.

Amplitude Spectrum The amplitude, particularly at lower frequencies, is strongly correlated with the image's overall style. This includes global characteristics like illumination, color balance, and contrast. High-frequency amplitudes, in contrast, relate more to fine-grained textures and noise.

FDA leverages this disentanglement in a simple yet effective manner. It is an untrained method that replaces the low-frequency amplitude spectrum of a source (synthetic) image with that of a randomly chosen target (real) image. The mathematical formulation is given in Eq. (1).

$$I_{s\to t} = \mathcal{F}^{-1}([\rho \odot |\mathcal{F}(I_t)| + (1-\rho)\odot |\mathcal{F}(I_s)|, \angle \mathcal{F}(I_s)]). \quad (1)$$

In Eq. (1), I_s and I_t are the source and target images, $\mathcal{F}$ is the Fourier transform, $|\cdot|$ and $\angle$ denote the amplitude and phase components, respectively. ρ is a binary mask that is 1 for a central, low-frequency region and 0 otherwise. This operation effectively pastes the low-frequency style of the real image onto the synthetic image while explicitly preserving the synthetic image's original phase.

3. Autonomous Spacecraft Temporal Rendezvous Datasets (ASTRA-HST)

Dataset Generation We constructed a custom simulator to produce high-fidelity synthetic image sequences of a chaser spacecraft approaching the Hubble Space Telescope (HST), chosen as the target for its complex geometry and publicly available NASA CAD model [1]. The simulator incorporates realistic orbital dynamics and lighting: the HST model (with solar panels and tubular body) is placed in a virtual 3D space environment with Earth albedo and direct

[1] https://science.nasa.gov/resource/hubble-space-telescope-3d-model/

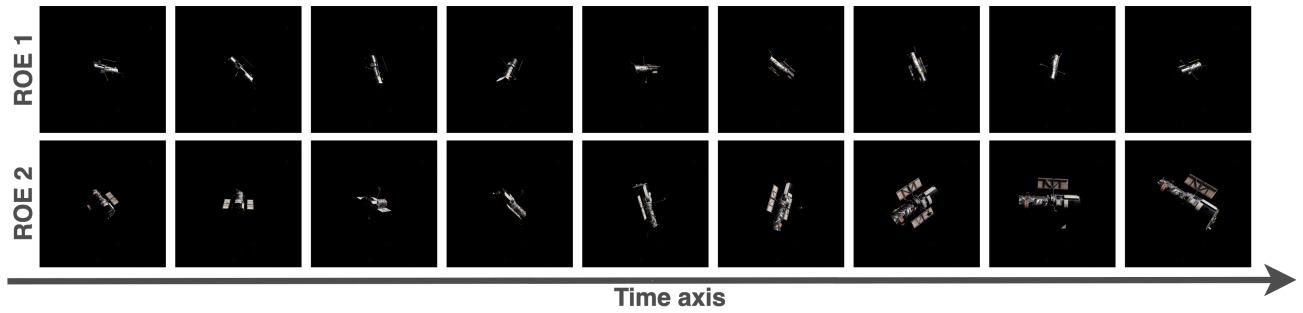


Figure 3. ASTRA-HST trajectory scenarios. (1) **ROE1 (v-bar hold scenario)** with near-stationary relative motion and constrained viewpoints; (2) **ROE2 (approach scenario)** with broader variation in distance.

sunlight illumination. We vary a comprehensive set of parameters for each simulated rendezvous scenario, including the chaser's initial position and velocity relative to HST, the HST's own rotation rate and axis, the approach trajectory profile (e.g. straight-line vs. curved approach), and the camera intrinsics and field-of-view. By sampling these parameters, we generated 960 distinct frames covering a broad envelope of conditions. Notably, as shown in Fig. 2, we use two camera FOV settings — a narrow $45°$ view and a wide $90°$ view — to simulate both inspection cameras and navigation cameras. For the $45°$ FOV, the relative distance between the chaser and HST ranges from approximately 70 m to 30 m, covering medium- to close-range inspection scenarios. For the $90°$ FOV, the distance spans 50 m down to 10 m, emphasizing closer-range navigation and docking conditions with stronger perspective distortion. The image resolution is 2048×2048 with realistic motion blur and sensor noise models applied. All sequences run at 24 frames per second, yielding smooth video streams; each sequence lasts several seconds, producing on the order of 960 frames (for a total of about 1000k). Every frame is time-stamped and labeled with the ground-truth 6-DoF pose of HST (relative to the camera) as well as a set of up to 36 visible 3D keypoints (landmarks on the HST model such as antenna tips, solar array corners, etc.). The annotation format follows the standards of prior datasets — we provide both continuous pose labels (i.e., quaternion and translation) and discrete keypoint coordinates in a JSON file per sequence, facilitating various learning approaches.

Furthermore, to enable the evaluation of sequential estimation algorithms such as filters and recurrent networks, we generated two distinct trajectory-based scenarios, following the methodology of prior work in satellite rendezvous [23]. For each trajectory, we provide 960 rendered frames; we release 256 trajectories per scenario (ROE1/ROE2), totaling 512 sequences. The scenarios are defined based on [10, 23]:

1) **ROE1:** A *v-bar hold* scenario (Fig. 3 (1)), where the chaser spacecraft maintains a fixed relative position along the target's velocity vector. This scenario exhibits limited motion and relatively constrained viewpoints, which makes it visually less diverse. While valuable for controlled evaluation, its restricted variability also renders test-time adaptation more challenging under domain gaps.

2) **ROE2:** An *approach* scenario (Fig. 3 (2)), where the chaser gradually closes in on the target spacecraft. The scenario generates a richer distribution of distances, viewing angles, and illumination conditions. This greater diversity supplies stronger adaptation signals, and in practice ROE2 has been observed to be more amenable to test-time adaptation than the more constrained ROE1 sequence.

This newly designed ASTRA-HST dataset, with its complex target shape and diverse, well-defined scenarios, provides a valuable resource for rigorously testing the limits of current pose estimation models and for developing the next generation of robust autonomous navigation systems.

Table 2. Summary of the ASTRA-HST dataset, highlighting its overall structure, annotation setting, and dataset scale.

Property	**Value / Description**
Sequential frame rate	24 fps
Number of frames (per trajectory)	960
Resolution (per frame)	2048×2048
Camera field of view (FOV)	$45°$(Standard), $90°$(Wide)
File formats	JPG, JSON (keypoints, 6-DoF)
Satellite model detail	>3.5 million polygons
Annotation types	3D keypoints map (up to 36)
Additional effects	Lens flare, Blooming, Vignetting
Total dataset size	~80 GB
Total trajectories	512
Total number of images	983,040

Diversity and Realism Our dataset introduces unprecedented diversity in sequential spacecraft imagery. In contrast to SPEED and SPEED+, which contain static images of a single spacecraft (Tango) under limited lighting variations, our sequences capture continuous relative motion between chaser and target, including realistic changes in scale (HST transitioning from far-field to near-field), occlusions

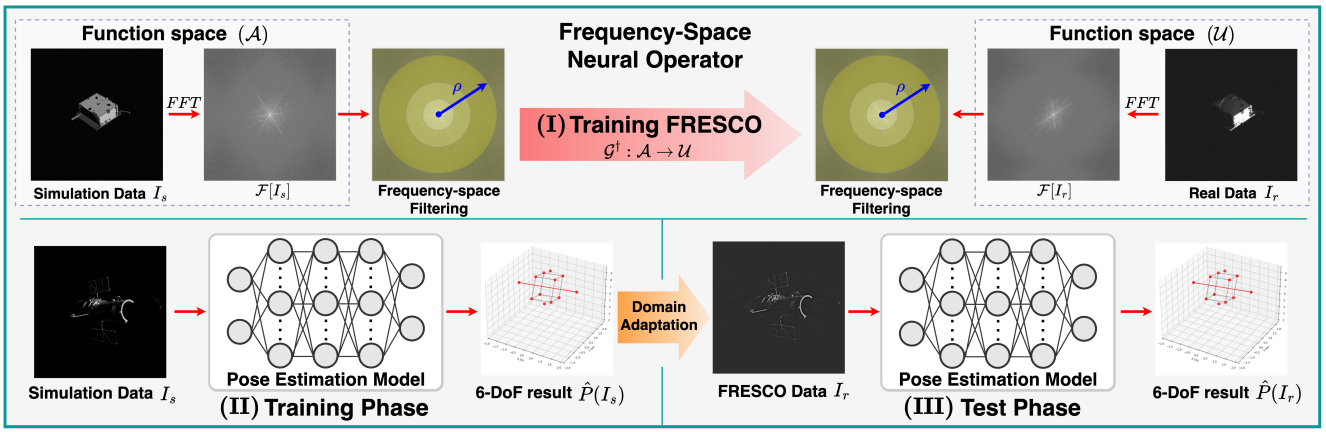


Figure 4. Overall pipeline: (I) FRESCO learns Fourier amplitude alignment between synthetic and real; (II) a pose regressor is trained on synthetic; (III) inference on FRESCO-transformed images with geometry preserved.

by HST's structure as viewing angle changes, and dynamic lighting (e.g. specular glints on rotating solar panels). Approach speeds range from slow drift to aggressive closing velocity, and HST's rotation covers single-axis spins as well as tumbling about multiple axes. We also randomize the initial orientation of HST in each scenario so that the trajectories explore varied attitudes (not always facing the camera). Table 2 summarizes key features of our ASTRA-HST dataset compared to existing public datasets for spacecraft pose estimation. Notably, ours is the first to offer hundreds of distinct trajectories (most prior datasets contain either no sequential data or only a couple of scripted trajectories). The field-of-view variation (45° vs. 90°) is another novel aspect, leading to substantial differences in image appearance. Wide-angle views capture more background and exhibit stronger perspective distortion, which poses additional challenges for pose estimation. For dataset generation, the 90° setting was rendered at closer ranges (10–50 m), since targets become too small at longer distances, whereas the 45° setting spans 30–70 m. The high resolution of our images (2048×2048) ensures that even at far distances, the target is clearly resolved, allowing accurate keypoint annotations across all frames.

4. Proposed Method: FRESCO

A natural starting point is frequency-domain style manipulation [4, 36, 38], which modifies low-frequency amplitude to transfer global appearance across domains. Building on this idea, we propose FRESCO, a neural operator that learns a band-selective mapping in the Fourier domain using paired synthetic–real satellite images from TRON. By learning the frequency-space transformation directly from data, FRESCO targets the stylistic discrepancy that drives the sim-to-real gap while preserving pose-critical structure.

4.1. Overall Pipeline

Fig. 4 summarizes our pipeline. In (I), both synthetic and real images are mapped to the frequency domain via FFT; a learnable amplitude-band filter then modifies only a selected region of the synthetic spectrum so that the transformed synthetic Fourier features align with the real distribution (phase is preserved). We subsequently train the pose regressor on synthetic images (II) and, at test time, feed FRESCO-transformed inputs (III) while predicting the same 6-DoF pose—the appearance becomes real-style, but geometry (and thus pose) remains encoded in the image.

4.2. Neural Operator Background

For learning mappings between function spaces, FRESCO draws upon the framework of Neural Operators (NO) [7, 11, 15, 16]. In particular, we adopt the philosophy of Fourier Neural Operator (FNO) [15], which operates directly in the frequency domain to approximate kernel integral operators efficiently. This makes FNOs a natural foundation for our frequency-space alignment, as they are well suited to modeling transformations across spectral bands while preserving geometric structure critical for pose estimation.

Let $\mathcal{A}$ and $\mathcal{U}$ be separable Banach spaces of functions defined on domain $\Omega \in \mathbb{R}^2$. A neural operator seeks to approximate an operator, which is defined as follows.

$$\mathcal{G}^\dagger : \mathcal{A} \to \mathcal{U}, \qquad (\mathcal{G}a)(x) = \int_\Omega \kappa(x,y)a(y)\,dy, \tag{2}$$

where $\kappa(x,y)$ is a learnable kernel. In practice, the operator is parameterized through iterative updates as follows.

$$v_{i+1}(x) = \sigma(W v_i(x) + (\mathcal{K}_\varphi v_i)(x)), \tag{3}$$

where W is a pointwise linear transform (e.g., a 1×1 convolution or skip-connection) , σ is a non-linear activation

function, and K_φ the kernel integral operator. Here, v_i denotes the intermediate representation at iteration i, i.e., a latent feature map lifted from the input.

In FNO, the kernel operator is efficiently implemented in Fourier space, as shown below.

$$(\mathcal{K}_\varphi v_i)(x) = \mathcal{F}^{-1}(R_\varphi(k) \cdot \mathcal{F}[v_i](k))(x), \tag{4}$$

where $\mathcal{F}$ denotes the Fourier transform, $R_\varphi(k)$ is a learnable Fourier filter, and k denotes the discrete frequency modes corresponding to the spatial coordinates.

4.3. Frequency-Space Mapping for Sim-to-Real

In our setting, images are treated as functions $I : \Omega \to \mathbb{R}^c$, where c is the number of channels. We denote by I_s a synthetic image and by I_r a real image. Applying the Fourier transform, we obtain the following.

$$\hat{I}_s(k) = \mathcal{F}[I_s](k), \qquad \hat{I}_r(k) = \mathcal{F}[I_r](k). \tag{5}$$

In Eq. (5), $\hat{I}(k)$ represents the complex Fourier coefficient at frequency mode k, with magnitude encoding amplitude and angle encoding phase. Our goal is to learn an operator $\mathcal{O}_\rho$ that reduces the spectral discrepancy between synthetic and real images while preserving pose-relevant geometric structure encoded in the phase as much as possible. To this end, we define $\mathcal{O}_\rho$ in the Fourier domain:

$$\hat{I}_{s\to r}(k) = \mathcal{O}_\rho(\hat{I}_s(k)) = R_\varphi(k;\rho) \cdot \hat{I}_s(k), \tag{6}$$

where $R_\varphi(k;\rho)$ is a learnable Fourier filter restricted to frequency bands within a radius ρ around the origin of the spectrum. As a result, this formulation ensures that:

- low- and mid-frequency amplitudes are selectively modified to inject realistic illumination and texture [20].
- phase components $\angle\hat{I}_{s\to r}(k) = \angle\hat{I}_s(k)$ remain unchanged, preserving pose-critical geometry.

Finally, the real-style image is obtained via inverse Fourier transform from $\hat{I}_{s\to r}(k)$:

$$I_{s\to r} = \mathcal{F}^{-1}\left[\hat{I}_{s\to r}(k)\right] = \mathcal{F}^{-1}[R_\varphi(k;\rho) \cdot \mathcal{F}[I_s](k)]. \tag{7}$$

Eqs (2)–(7) establish a natural connection between the classical FNO formulation and our proposed FRESCO operator. While FNO parameterizes a global kernel $R_\varphi(k)$ to approximate PDE solution operators, FRESCO specializes this idea by introducing a band-selective Fourier operator $\mathcal{O}_\rho$ that learns to transfer spectral style between synthetic and real images.

5. Experiments

In this section, we compare the performance of existing baseline models on our proposed dataset. The ASTRA-HST dataset is split at the trajectory level into training, validation, and test sets with a ratio of 0.7/0.1/0.2. Only the synthetic training and validation images were used for model training, while evaluation is conducted on both synthetic and FRESCO-generated test images. For completeness, we also evaluate on the SHIRT dataset using models trained on the publicly available SPEED+ synthetic data. Our software and hardware environments are as follows: UBUNTU 20.04 LTS, PYTHON 3.12, PYTORCH 2.7, CUDA 12.2, and NVIDIA RTX A6000 GPUs.

5.1. Baselines

In our experiments, we use five representative baseline models for satellite pose estimation tasks on our proposed ASTRA-HST dataset.

U-Net [30]: The classical encoder–decoder convolutional network originally proposed for biomedical image segmentation has been widely adopted as a generic backbone for dense prediction tasks. In our baseline setup, U-Net is adapted to spacecraft imagery for pixel-wise regression tasks such as keypoint heatmap prediction.

KRN [26]: The Keypoint Regression Network is a lightweight CNN designed for single-image spacecraft pose estimation. It regresses the 2D image locations of a predefined set of surface keypoints. The full 6-DoF pose is then recovered via a Perspective-n-Point (PnP) solver [14, 18] using the known 3D coordinates of these keypoints.

SPNv2 [25]: The Spacecraft Pose Network v2 is a multi-task CNN developed to address domain gaps in spacecraft pose estimation. It integrates a shared encoder with multiple prediction heads, including bounding box detection, direct pose regression, heatmap-based keypoint localization, and binary segmentation of the spacecraft foreground.

SPNv3 [22]: The Spacecraft Pose Network v3 leverages Vision Transformer (ViT) [2] backbones to balance computational efficiency and robustness across the synthetic-to-real gap. Extensive design-space exploration revealed that ViT architectures, combined with strong data augmentation, achieve state-of-the-art robustness on HIL datasets such as SPEED+ while maintaining flight-ready efficiency.

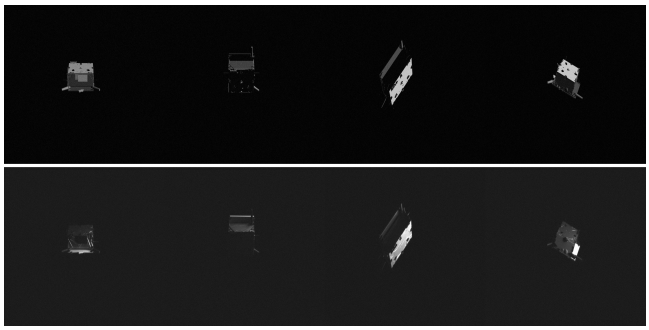
AUKF [24]: An online supervised training method that leverages an onboard Kalman filter [6, 35] to generate pseudo-labels from flight images, enabling continual adaptation of the pose network during proximity operations.

5.2. Evaluation Metrics

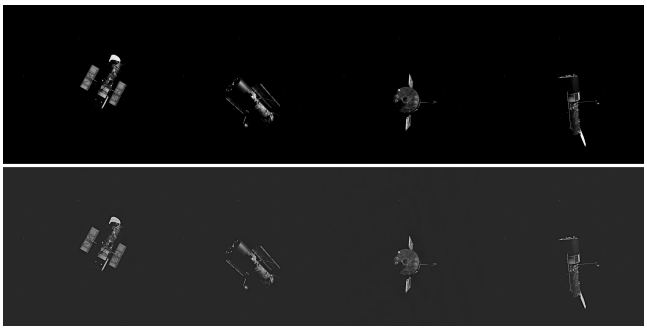
Translation Error (E_t) This metric measures the absolute difference between the predicted and ground-truth positions. It is calculated as the Euclidean (L_2) norm of the vector difference, expressed in meters:

$$E_t = ||\hat{t} - t||_2, \tag{8}$$

where $\hat{t}$ is the predicted vector representing the estimated 3D position of the target's origin in the camera's coordinate

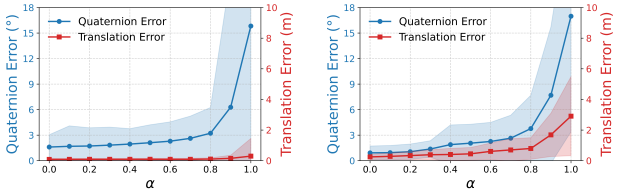


(a) SHIRT: synthetic (top) vs. TRON (bottom).



(b) ASTRA-HST: synthetic (top) vs. FRESCO (bottom).

Figure 5. Visual comparison of synthetic and target-domain appearance. Panel (a) shows SHIRT with paired synthetic and TRON real images; panel (b) shows ASTRA-HST with synthetic inputs and FRESCO outputs in a real-style, sensor-consistent appearance. In both panels the top row is synthetic and the bottom row is the target domain.



(a) SHIRT (b) ASTRA-HST

Figure 6. **Appearance interpolation results.** Pose errors under interpolation from synthetic ($\alpha = 0$) to target domain ($\alpha = 1$). (a) SHIRT with TRON real targets; (b) ASTRA-HST with FRESCO real-style targets.

frame, and t is the ground-truth translation vector.

Rotation Error (E_R) This metric quantifies the angular distance between the predicted and ground-truth orientations. It is calculated as the angle of the error quaternion, which represents the rotation needed to align the predicted orientation with the true orientation. Given as twice the arccosine of the absolute value of the inner product of the two unit quaternions, the error is expressed in degrees:

$$E_R = 2 \cdot arccos(| < \hat{q}, q > |), \tag{9}$$

where $\hat{q}$ is the predicted rotation represented as a 4D unit quaternion, and q is the ground-truth rotation, also represented as a 4D unit quaternion. $< \cdot, \cdot >$ is the inner product of the two quaternions, and $| \cdot |$ denotes the absolute value of the scalar result.

Pose Error (E_{pose}) To provide a single, unified score that balances both translation and rotation errors, we define E_{pose} as follows.

$$E_{pose} = E_R \text{ [rad]} + \frac{E_t}{||t||_2}. \tag{10}$$

This score is the sum of the orientation error (in radians) and the translation error normalized by the magnitude of the ground-truth distance. This normalization penalizes position errors more heavily at closer ranges, which is critical for proximity operations.

5.3. Analysis of Baseline Results

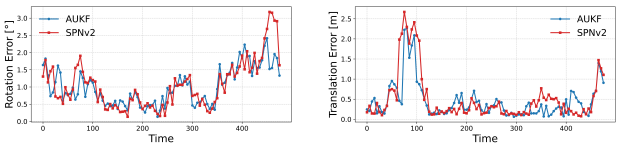
Table 3 shows the comparative performance of all models under synthetic and FRESCO settings for both $45°$ and $90°$ FOVs. Across all four architectures, the narrow FOV $45°$ synthetic setting yields consistently lower pose errors than the wide FOV $90°$ synthetic setting, with the gap driven primarily by rotation error. For example, SPNv2 and SPNv3 roughly double their rotation error when moving from $45°$ to $90°$, which in turn inflates the E_{pose}. This trend is consistent with the dataset design: the $90°$ camera introduces stronger perspective distortion and broader context, which tends to exacerbate foreshortening, background clutter, and occlusion effects that particularly harm orientation recovery. In contrast, the $45°$ view provides more stable local geometry and fewer extreme viewpoints, making pose inference easier despite the greater average distance.

When we replace synthetic inputs with FRESCO real-style images, we observe a consistent degradation in accuracy (higher E_{pose}) across most models and both FOVs (see Table 3). This indicates that the appearance shift induced by FRESCO—while preserving geometry—is sufficient to expose a non-trivial sim-to-real gap on ASTRA-HST, providing an HIL-free and controllable way to probe robustness.

As shown in Fig. 7, applying the Adaptive Unscented Kalman Filter (AUKF) on top of SPNv2 significantly mitigates extreme error spikes in both rotation and translation, yielding more stable sequential predictions. This improvement highlights the benefit of integrating sequential filtering into pose estimation pipelines. However, the persistence of non-trivial error regions indicates that the ASTRA-HST sequential dataset poses substantial challenges that are not yet

Table 3. Comparison of baseline models on ASTRA-HST across synthetic and FRESCO-generated inputs at two camera FOV settings (45° and 90°). Lower values indicate better performance for translation error E_t, rotation error E_R, pose error E_{pose}.

Model	Synthetic (FOV 45°)			Synthetic (FOV 90°)			FRESCO-generated (FOV 45°)			FRESCO-generated (FOV 90°)		
	E_t[m]	E_R[°]	E_{pose}[-]	E_t[m]	E_R[°]	E_{pose}[-]	E_t[m]	E_R[°]	E_{pose}[-]	E_t[m]	E_R[°]	E_{pose}[-]
U-Net [30]	3.42	32.65	0.65	3.24	42.05	0.87	3.46	35.60	0.67	3.45	42.97	0.88
KRN [26]	3.50	42.03	0.84	3.52	49.02	0.98	3.55	49.22	0.99	3.52	56.07	1.21
SPNv2 [25]	0.82	4.19	0.09	0.88	9.32	0.19	0.84	6.51	0.13	0.86	13.19	0.19
SPNv3 [22]	0.79	6.75	0.13	0.69	11.21	0.22	0.81	8.06	0.16	0.82	13.30	0.20



(a) Rotation error comparison (b) Translation error comparison

Figure 7. Error curves on the ASTRA-HST sequential test set, evaluated on FRESCO-generated real-style images, comparing SPNv2 alone against SPNv2 combined with the AUKF filtering.

fully addressed by existing methods. This observation supports the claim that our sequential benchmark is of long-term research value, providing a fertile testbed for advancing both filtering techniques and domain-robust pose estimation architectures.

5.4. Domain Shift: Qualitative Sim-to-Real Comparison and Domain-Interpolation Stress Test

This subsection connects qualitative Sim-to-Real examples presented in Fig. 5 with a quantitative domain-interpolation stress test.

Qualitative domain comparison (Sim-to-Real) Fig. 5 contrasts (top) synthetic and (bottom) real images for SHIRT, and synthetic inputs with FRESCO real-style outputs for ASTRA-HST. Crucially, FRESCO is trained on SHIRT synthetic–real pairs and then applied to ASTRA-HST synthetic frames; the resulting real-style outputs are qualitatively in line with SHIRT real imagery—i.e., they adopt a more real-like appearance—while preserving pose-critical geometry (e.g., edges and keypoint layout). This cross-dataset visual transfer suggests that FRESCO effectively carries the SHIRT real appearance to HST, motivating the domain-interpolation stress test that follows.

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

$$x(\alpha) \;=\; (1-\alpha)\, x_{\text{syn}} \;+\; \alpha\, x_{\text{real}}, \qquad \alpha \in [0,1], \quad (11)$$

where x_{syn} is the synthetic rendering and x_{real} denotes the real image (TRON, for SHIRT) or the FRESCO real-style image (for ASTRA-HST).

For evaluation, we first identify the 30 most challenging frames at the real endpoint ($\alpha = 1$), i.e., those with the highest baseline errors. We then trace their performance as α increases from 0 to 1. Fig. 6 reports mean and standard deviation of both rotation error (degrees, left axis) and translation error (meters, right axis).

Across both datasets, errors remain low in the synthetic regime but rise sharply as $\alpha \to 1$, revealing a clear *loss barrier* [17] where small domain changes produce disproportionately large error increases. Crucially, this barrier emerges not only with SHIRT real images but also with ASTRA-HST FRESCO real-style images, indicating that FRESCO reproduces the failure signature observed on real data and can serve as a practical proxy for robustness diagnosis when paired real imagery is limited.

6. CONCLUSIONS

To support systematic evaluation for spacecraft pose estimation, we release a new sequential HST dataset comprising two trajectory scenarios—*v-bar hold* and *approach*—at multiple fields of view and under diverse illumination, background, and sensor conditions. Each sequence is annotated with full 6-DoF pose and keypoint correspondences, enabling both frame-level and sequence-level benchmarking and facilitating studies on domain shift and sequential filtering for robust navigation.

We also introduce FRESCO, which learns band-specific mappings in the Fourier domain to adjust global illumination and fine textures while preserving pose-critical geometry. By operating on amplitude spectra while keeping phase fixed, FRESCO provides a principled and scalable mechanism to bridge the appearance gap between synthetic renderings and real-style imagery. Our experiments focus on translating synthetic imagery toward publicly available real-style imagery to emulate on-orbit appearance without relying on a physical HIL facility. Future work will extend FRESCO to genuine on-orbit imagery, incorporate physics-aware priors, and evaluate downstream pose networks within closed-loop navigation pipelines, with emphasis on long-duration tracking stability and robust autonomous rendezvous.

References

[1] 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*, pages 2781–2790, 2022. 1

[2] 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. 6

[3] 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. 1, 2

[4] 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*, pages 6891–6902, 2021. 5

[5] Abhishek Kadian, Joanne Truong, Aaron Gokaslan, Alexander Clegg, Erik Wijmans, Stefan Lee, Manolis Savva, Sonia Chernova, and Dhruv Batra. Sim2real predictivity: Does evaluation in simulation predict real-world performance? *IEEE Robotics and Automation Letters*, 5(4):6670–6677, 2020. 1

[6] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960. 6

[7] 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. 5

[8] 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*, pages 2938–2946, 2015. 1

[9] 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. 1, 2

[10] 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. 4

[11] 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. 5

[12] Rahul G Krishnan, Uri Shalit, and David Sontag. Deep kalman filters. *arXiv preprint arXiv:1511.05121*, 2015. 3

[13] 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*, pages 21285–21295, 2023. 3

[14] 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. 6

[15] 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. 5

[16] 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. 5

[17] 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. 8

[18] 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. 6

[19] 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. 2

[20] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 2005. 6

[21] 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. 1

[22] Tae Ha Park and Simone D'Amico. Bridging domain gap for flight-ready spaceborne vision. *arXiv preprint arXiv:2409.11661*, 2024. 2, 6, 8

[23] 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. 2, 4

[24] 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)*, pages 11744–11752. IEEE, 2024. 3, 6

[25] 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, 2024. 2, 6, 8

[26] Tae Ha Park, Sumant Sharma, and Simone D'Amico. Towards robust learning-based pose estimation of noncooperative spacecraft. *arXiv preprint arXiv:1909.00392*, 2019. 6, 8

[27] 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)*, pages 1–15. IEEE, 2022. 1, 2

[28] 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. 2

[29] 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. 3

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 6, 8

[31] Sumant Sharma, Tae Ha Park, and Simone D'Amico. Spacecraft pose estimation dataset (speed). *Stanford Digital Repository*, 2019. 1

[32] Long Tian, Changjae Oh, and Andrea Cavallaro. Test-time adaptation for 6d pose tracking. *Pattern Recognition*, 152: 110390, 2024. 3

[33] 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)*, pages 23–30. IEEE, 2017. 1, 3

[34] Szabolcs Velkei, Csaba Goldschmidt, and Károly Vass. A large-scale, physically-based synthetic dataset for satellite pose estimation. *arXiv preprint arXiv:2506.12782*, 2025. 3

[35] Greg Welch, Gary Bishop, et al. An introduction to the kalman filter. 1995. 6

[36] 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*, pages 14383–14392, 2021. 5

[37] 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*, pages 14880–14890, 2022. 3

[38] 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*, pages 4085–4095, 2020. 3, 5