

Article

From Sim to 6DOF: Deep Learning for Real-Time Satellite Pose Estimation from Resolved Ground-Based Imagery

Thomas Dickinson ¹, Dawson Friesenhahn ², Justin Fletcher ², Derek Walvoord ¹, Dennis Montera ² and Michael Gartley ^{1,*} 

¹ Chester F. Carlson Center for Imaging Science, Rochester, NY 14623, USA; twd5708@rit.edu (T.D.); djwcis@rit.edu (D.W.)

² Air Force Research Laboratory, Kirtland Air Force Base, Albuquerque, NM 87123, USA; dawson.friesenhahn.1@spaceforce.mil (D.F.); justin.fletcher.14.ctr@spaceforce.mil (J.F.)

* Correspondence: mxgpci@rit.edu

Abstract

This work presents the first complete system for automated six degrees of freedom (6DOF) satellite pose estimation from spatially resolved, ground-based, adaptive optics (AO)-corrected imagery, addressing a key challenge in Space Domain Awareness (SDA). The approach mitigates the need for human labeling by directly regressing satellite orientation and position from blurry, noisy, and deeply shadowed imagery. A multi-stage deep neural network pipeline localizes the satellite, predicts pose, and optionally applies temporal filtering. Networks are trained exclusively on fully synthetic imagery generated from a CAD model, yet generalize effectively to real data, bridging the Sim2Real domain gap. On 137 real, human-labeled test images of Seasat, the model achieved a mean rotation error of 5° and a mean image-plane translation error of 21 cm. Slant range error was quantitatively evaluated on synthetic data due to unknown real-sensor parameters. Qualitative evaluation of additional real Seasat imagery rated 177 of 199 predicted poses as “ground truth equivalent” or “high-confidence match,” with zero catastrophic failures. The system was extended to seven degrees of freedom (7DOF) for satellites with articulating components and demonstrated on real Hubble Space Telescope (HST) imagery, achieving 5.5° rotation error, 51 cm image-plane translation error, and 8° symmetry-adjusted solar array error on a 249-frame pass with causal temporal filtering. Across 586 real test images from Seasat and HST (captured over multiple decades under diverse conditions) the system consistently performed well. Full 6DOF performance was quantified on a high-fidelity wave optics (HFWO) synthetic test set of Seasat, where the model achieved 8.4° mean rotation error, 34 cm image-plane translation error, and 1.4% line-of-sight range error at $r_0 = 6$ cm and 1031 km range. In a limited 200-image benchmark, the model demonstrated 48% lower mean rotation error than a single human labeler while operating $\sim 800\times$ faster. It required <40 h and a single A100 GPU to generate data and train. The approach was also demonstrated for ARGOS, a smaller satellite with highly symmetric geometry. An exploratory General Image-Quality Equation-based image quality metric (AO-IQ) was introduced as an empirical correlate for pose accuracy. General-purpose models like GPT-4o and Depth Anything V2 failed across most SDA tasks, but rapid gains in vision-language models warrant continued monitoring. These results establish a new operational baseline for practical, real-time satellite pose estimation from AO SDA imagery.



Academic Editor: Vladimir S. Aslanov

Received: 30 June 2026

Revised: 13 August 2026

Accepted: 17 August 2026

Published: 19 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

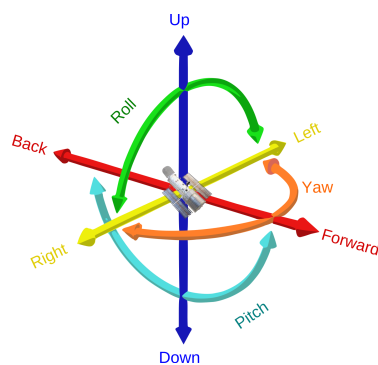
[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: satellite pose estimation; adaptive optics; Space Domain Awareness; sim2real; domain gap; deep learning; 6DOF; n-DOF

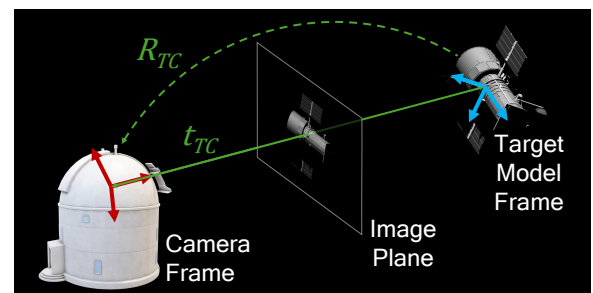
1. Introduction

Automated satellite pose estimation has the potential to enhance spacecraft health assessment, pattern-of-life analysis, and space traffic management, becoming an integral component of future Space Domain Awareness (SDA) architectures. Large-aperture, ground-based Electro-Optical (EO) instruments routinely collect adaptive optics (AO)-corrected, spatially resolved imagery of low Earth orbit (LEO) satellites [1]. In practice, difficulties arising from visual ambiguity and labor-intensive workflows mean that few of these images are used for pose estimation. Automating pose estimation from such imagery would enable timely characterization of satellite behavior, thereby increasing the operational value of mission-relevant AO assets. Additionally, the rapidly growing population of objects in LEO motivates increased automation.

This work addresses monocular six degrees of freedom (6DOF) pose estimation, the recovery of both three degrees of freedom (3DOF) position and 3DOF orientation from a single image. 6DOF pose is illustrated in Figure 1a, and 6DOF pose estimation is illustrated in Figure 1b. In target platforms featuring articulating components (such as the solar arrays on the Hubble Space Telescope (HST) that track the Sun with one rotational degree of freedom) estimating additional internal parameters raises the overall problem dimension ($\geq 6\text{DOF}$). The problem is tractable for SDA because image data already exists, but exploitation remains technically challenging under operational conditions: partial illumination with deep shadows, low and variable signal-to-noise ratio, poor contrast, residual blur that varies with atmospheric conditions (r_0) and AO-system performance, and wide target slant range variation. While supervised deep learning can address these visual degradations (illustrated in Figure 2), labeled real imagery is virtually non-existent. Consequently, models must be trained entirely in simulation and bridge the Sim2Real domain gap.



(a) Six degrees of freedom.



(b) Coordinate system transform.

Figure 1. (a) The 6DOF for rigid object pose: translation along and rotation about three orthogonal axes (figure adapted under CC BY-SA 4.0 license from [2]). (b) Satellite pose estimation determines the translation vector t_{TC} and rotation matrix R_{TC} from the satellite (blue axes) to the camera (red axes); the image plane encodes the observed 2D projection used to infer this transformation.

Two precursors explored satellite pose estimation from such imagery. Wood [3] introduced a matched filter-based approach that demonstrated classification of discrete two degrees of freedom (2DOF) rotations using synthetic imagery with full frontal illumination and fixed image-plane translation and target range, an important starting point that did not address full 6DOF pose and lacked validation on real data. Lucas et al. [4] presented the first end-to-end deep-learned approach, predicting 3DOF rotation using quaternion regression on synthetic AO imagery, showing promising results on pristine renders, but degraded accuracy with realistic residual AO blur (uncorrected atmospheric distortion). Image-plane translation and target range were fixed and not estimated, the renders were fully front

illuminated, and no real imagery was tested. These studies established the problem, highlighted core challenges, and demonstrated key techniques and advancements but did not deliver a practical capability suitable for operational SDA.

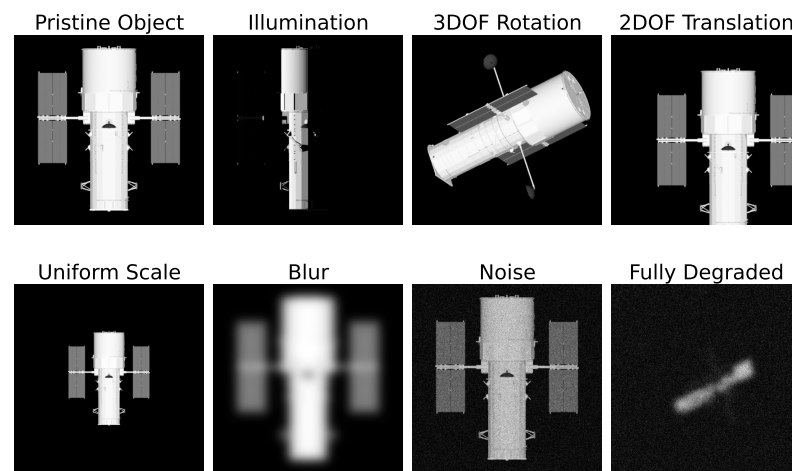


Figure 2. A pristine render of a Hubble Space Telescope (HST) Computer-Aided Design (CAD) model is used to independently illustrate the visual confusion resulting from partial illumination, 3DOF rotation, 3DOF translation, blur, and noise. While HST’s pose in the top left image is obvious, the combined effects of these transformations result in a much more ambiguous pose in the final image.

Progress in direct pose regression has been driven by two main developments that inform this work. First, GDR-Net [5] and its predecessor CDPN [6] showed that semantic-segmentation-guided supervision with dense pixelwise $2D \leftrightarrow 3D$ correspondences and a learned Perspective-n-Point (PnP)-inspired regression head can effectively replace classical PnP /Random Sample Consensus (RANSAC). This approach elegantly simplifies the learning of the continuous output-space pose regression task by way of a semantic segmentation geometric classification task, which has a discrete output space that has arguably better learnable mappings and reduced loss surface complexity. These works also introduced and employed Scale-Invariant Representation for Translation Estimation (SITE), which is adopted and adapted for this work. Subsequent variants of GDR-Net, including GDR-NPP [7] and GPose2023 [8], improved robustness and won the 6D Object Localization tasks of the Benchmark Object Pose (BOP) 2022 and 2023 challenges [9,10]. Second, the continuous 6D representation for 3DOF rotation (Ref. [11], employed in GDR-Net and subsequent variants) eliminates discontinuities present in other 3DOF representations (like quaternions) and stabilizes training for direct pose regression. The representation allows incorporation of Gram-Schmidt orthogonalization as a layer within the regression head, enforcing learning and output of physically valid 3DOF rotations (orthonormal rotation matrices, special orthogonal group in 3D ($SO(3)$)). This representation paired with in-network orthogonalization is now standard in leading pose regression models and is adopted here for our application.

Lessons from the Satellite Pose Estimation Challenge (SPEC) further shape the approach. Satellite Pose Estimation Challenge 2019 (SPEC2019) [12] established the strength of sparse keypoint correspondences and PnP /RANSAC pipelines on synthetic Rendezvous and Proximity Operations (RPO) imagery, while Satellite Pose Estimation Challenge 2021 (SPEC2021) [13] emphasized bridging the Sim2Real domain gap using challenging and highly realistic Hardware-In-the-Loop (HIL) test images and synthetic training images. Park et al. [14] introduced SPNv2 with multi-task learning, the 6D rotation representation, and domain-specific randomized augmentations that target realistic illumination, blur,

and noise distributions. They later added temporal processing and improved efficiency for edge deployment [14–16]. Their approaches for randomized training data augmentation strongly inspired this work. Although SPEC was limited to 6DOF, did not provide a Computer-Aided Design (CAD) model, and used high-resolution space-based RPO imagery rather than lower-quality ground-based AO imagery, the challenge and resulting publications provided a wealth of knowledge to inform the model architecture, data generation, and augmentation choices employed here.

To systematically evaluate the state of the art and contextualize the technical advancements of this work, Table 1 compares the present pipeline against four key precursors: Wood (1996) [3], Lucas et al. (2020) [4], GDR-Net (2021) [5], and SPNv2 (2024) [14].

Table 1. Systematic comparison of pose estimation pipelines across platforms, image quality, estimated DOF, validation scope, temporal processing, and inference rate.

Method	Platform	Image Quality	DOF	Validation	Temporal	Inference Rate
Wood [3]	Ground AO	AO blur, noise	2DOF ^a	Syn.	None	N/A
Lucas et al. [4]	Ground AO	AO blur, noise	3DOF ^a	Syn.	None	N/A
GDR-Net [5]	General Purpose	High res. RGB	6DOF	Syn.+Real	None	~45 Hz ^b
SPNv2 [14]	Space RPO	High res.	6DOF	Syn.+HIL ^c	KF	~1.6 Hz ^d
Present Work	Ground AO	AO blur, noise	≥6DOF ^e	Syn.+Real ^e	KF	7.1 Hz ^f

^a Fixed target translation/range and full frontal illumination. ^b Evaluated on a desktop GPU (NVIDIA RTX 2080 Ti; NVIDIA Corporation, Santa Clara, CA, USA). ^c Evaluated on synthetic images, real images captured in a lab: Hardware-In-the-Loop (HIL), and 25 real flight images. Notably, range error was not presented for the 25 real flight images. ^d Evaluated on a desktop CPU (Intel i9-9900K @ 3.60 GHz; Intel Corporation, Santa Clara, CA, USA), trained on four NVIDIA V100 GPUs; NVIDIA Corporation, Santa Clara, CA, USA. ^e Full ≥ 6DOF (including solar array articulation) is quantitatively validated on synthetic data; real imagery quantitatively validates 3DOF attitude + 2DOF image-plane position + 1DOF articulation with range estimates input to model and range accuracy unevaluated due to unknown sensor IFOV and ground-truth range. ^f Inference rate measured on a consumer laptop GPU (NVIDIA RTX 3060 Mobile; NVIDIA Corporation, Santa Clara, CA, USA). Data generation and model training require 40 h on a single NVIDIA A100 GPU (NVIDIA Corporation, Santa Clara, CA, USA).

To resolve the Sim2Real operational gap identified in prior ground-based AO efforts [3,4], this study presents a practical three-stage pipeline consisting of a custom Coarse Localizer for Region of Interest (RoI) cropping, a direct-regression PoseNet utilizing the 6D rotation representation [11], and an adaptive Kalman filter (AKF) for temporal filtering. The system assumes a single target instance in the field of view and an available 3D CAD model at train time. To improve accuracy when operationally deployed, we provide a two-line element set (TLE)-assisted pipeline to partially address scale ambiguities. At inference time, the network ingests a prior physical scale estimate computed from the satellite's TLE or Orbit Mean-Elements Message (OMM) ephemeris paired with the sensor's Instantaneous Field of View (IFOV). This prior bounds the sensor-target slant range (Z), while image-space translation (X, Y) is constrained to keep the target within the sensor's field of view. Only the 3DOF rotation is completely unconstrained.

Because public AO image sequences lack ground-truth sensor IFOV and slant range, real-data quantitative evaluations in this work assess five degrees of freedom (5DOF) spatial pose (consisting of 3DOF orientation and 2DOF image-plane translation) alongside one degree of freedom (1DOF) articulation when applicable. Full 6DOF/seven degrees of freedom (7DOF) pose performance (including slant range estimation accuracy) is quantitatively evaluated on high-fidelity synthetic test sets. Uniquely, this approach enhances real-time sensor operator situational awareness by making unobservable, unilluminated target ge-

ometry immediately visible, as demonstrated in Figure 3. This operational capability is enabled through distinct contributions across three primary areas:

- **System Integration and Operational Pipeline:** Development of an end-to-end framework operating in real time (>7 Hz) trained exclusively on synthetic CAD renders that generalizes to a dataset of 586 real images across two distinct satellite targets (Seasat and HST) captured over decades by two independent AO facilities in New Mexico and Hawaii.
- **Architectural Adaptations:** A serial backbone architecture of four compact sequential U-Nets (reconstruction $\rightarrow$ relighting $\rightarrow$ surface-region classification $\rightarrow$ dense 3D point regression) that provides dense intermediate supervision prior to pose regression. By leveraging synthetic ground-truth labels at each intermediate stage, the model incrementally learns rich geometric representations that ultimately contribute to pose accuracy. This serial architecture enables training on a single A100 graphics processing unit (GPU) (<40 h including data generation) and real-time inference on a consumer laptop.
- **Application-Specific Adaptations:** Integration of physical priors (TLE-derived scale representation and image-space illumination vectors), SITE-inspired scale normalization reparameterized for megameter-scale slant ranges, and a lightweight empirical domain-specific augmentation pipeline simulating ground-based AO degradations (core-halo double-Gaussian point spread functions (PSFs), specular glints). Crucially, this demonstrates that parametric augmentations can bridge the Sim2Real gap without requiring computationally prohibitive wave-optics simulations at train time.

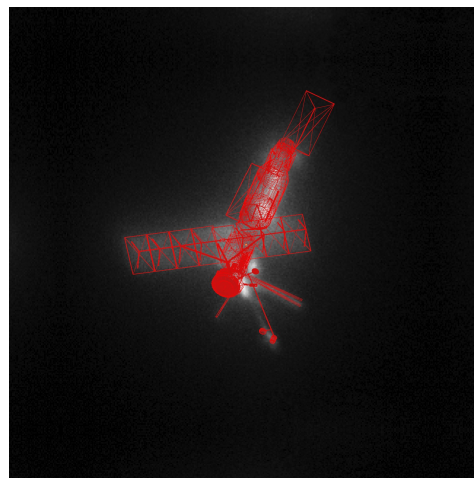


Figure 3. A 6DOF pose prediction was produced for the underlying grayscale image of the satellite Seasat (SATCAT #10967). The pose prediction was used to align a CAD model, and the model wireframe was overlaid in red on the grayscale image. In this example the previously unobservable position of the unilluminated solar arrays becomes apparent to the viewer.

Additionally, this work quantifies how pose accuracy correlates with atmospheric parameters (r_0) and image quality (adaptive optics image quality (AO-IQ)), benchmarks performance of a single human pose labeler, and evaluates contemporary Vision-Language Models (VLMs) on pose estimation tasks. Preliminary results were published in the proceedings of the 2024 AMOS Conference [17], and additional methodological details and extended analyses are provided in [18].

2. Materials and Methods

2.1. Approach Overview

Large-scale synthetic data are generated with Blender/EEVEE using site-specific terminator illumination, scaled orthographic rendering, and AO-PSF inspired degradations based on an empirical 2D double-Gaussian PSF [19]. A learned end-to-end pipeline performs coarse amodal localization followed by 6DOF pose regression (extendable to n -DOF by applying $n - 6$ main body pose-aware branched regression heads). Optionally, temporal filtering is applied when time-series data are available. Besides the grayscale image, two application-specific inputs are injected throughout: an illumination direction unit 3-vector in image coordinates and an orthographic scale estimate (object-plane horizontal field of view (HFOV), a linearly scaled proxy for target range). Model architectures are shown in Figures 4 and 5 and implemented entirely in TensorFlow [20]. High Fidelity Wave Optics (HFOW) closed-loop AO-corrected imagery was simulated for test-time evaluation (Section 2.4).

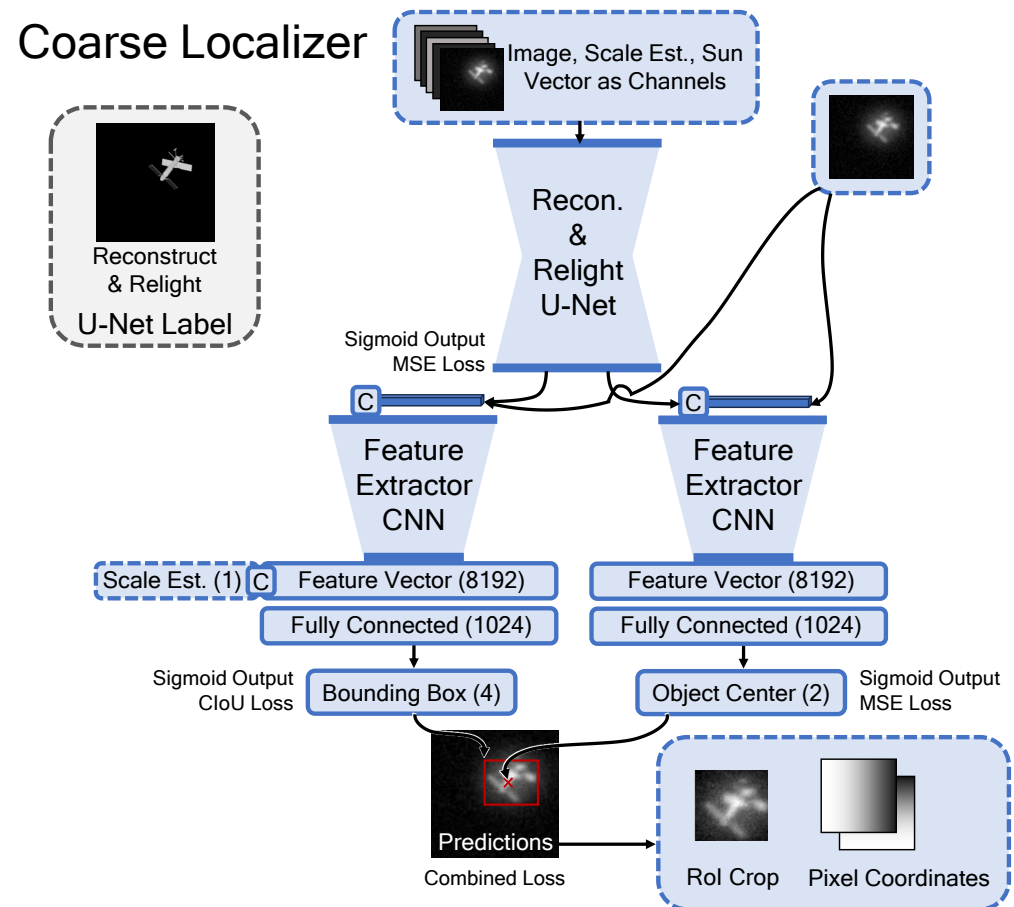


Figure 4. Architecture of the Coarse Localizer. The Coarse Localizer takes in 256×256 pixel grayscale images, illumination vectors, and normalized orthographic scale estimates (object plane HFOV in meters). It first uses a U-Net backbone to regress an intermediate reconstructed and relit image which is then fed (along with the initial input image) into two separate convolutional neural network (CNN) heads to output bounding box coordinates and object center coordinates. These outputs are used to crop and resize the input image to a 64×64 pixel RoI and to generate p_{xy} pixel coordinate maps.

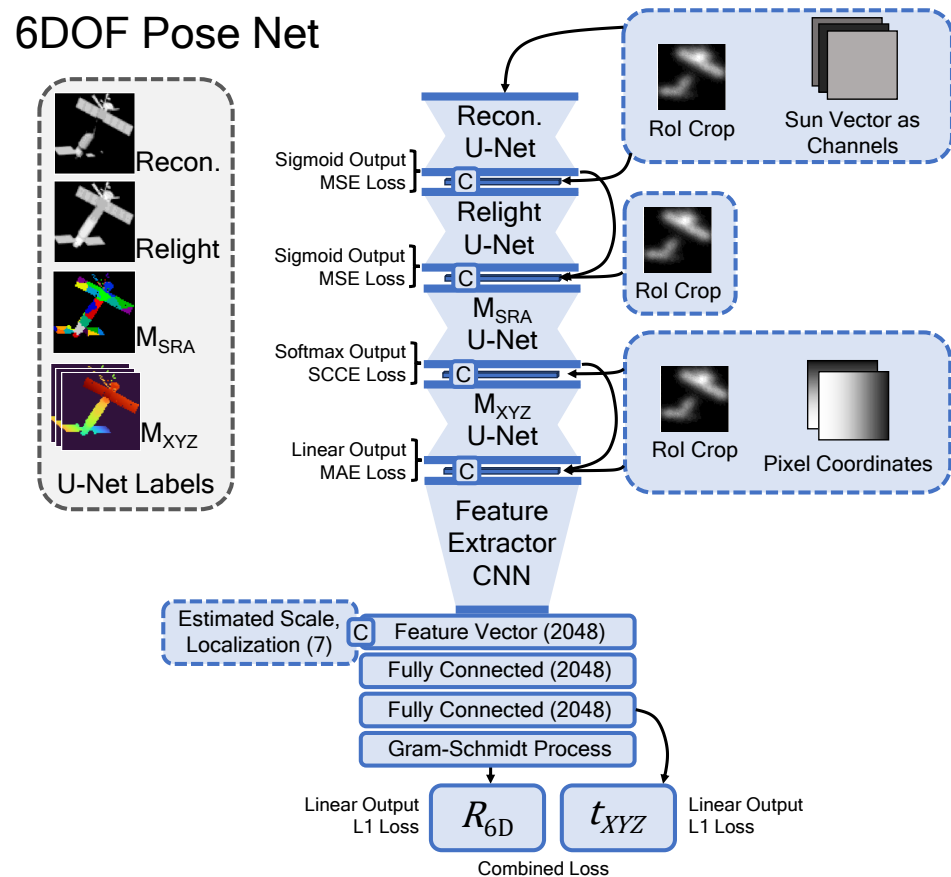


Figure 5. Architecture diagram for the PoseNet. The PoseNet takes in the RoI image, the p_{xy} pixel coordinate maps, an orthographic scale estimate (a representation of target slant range that is calculated based on known camera intrinsic parameters and a TLE), the predicted bounding box coordinates, and the predicted object center coordinates. The PoseNet predicts 6DOF pose, with surface region attention maps and 3D point maps being intermediate outputs.

2.2. Imaging Geometry, Coordinates, and Camera Model

Topocentric azimuth and elevation for the Sun are computed with Astropy across a full year and restricted to morning/evening terminators (three hours before sunrise to sunrise, sunset to three hours after sunset) [21]. These coordinates are perturbed by additive noise sampled from a normal distribution, $\mathcal{N}(0^\circ, 1^\circ)$, to augment the dataset. Samples for which the Sun-target vector intersects the Earth were excluded using a spherical Earth model with a radius of $R_E = 6357$ km, chosen to prevent erroneously discarding valid terminator geometries (Figure 6a). The range of Sun positions present during terminator conditions at a fixed site varies by $<0.1^\circ$ over 20 years, supporting trained model reuse across years but not across ground sites. Target azimuth angles were sampled uniformly (Figure 6b), while zenith angles were drawn from a uniform distribution U on $[0, 1]$ with a power-law transformation with exponent 0.4, i.e., $\theta_z = 60^\circ \cdot U^{0.4}$. This biases the resulting elevations, $90^\circ - \theta_z$, toward lower values (Figure 7), with the constraint that target elevation angles exceed 30° .

Topocentric Sun and target distributions for training data are displayed in Figure 6. The render camera employs scaled orthographic projection at 256×256 px. The IFOV is selected so that the target's longest axis subtends $\leq 50\%$ of the image HFOV at minimum range; for Seasat this yields 139.2 nrad. Alternatively, the IFOV can be selected to equal that of a real system or to be a multiple of a real system's IFOV. This was done for HST, where simulation IFOV $\simeq 200$ nrad was selected as double the true system IFOV. Blender axes follow +X right, +Y up, +Z out of the image plane (sensor-to-target).

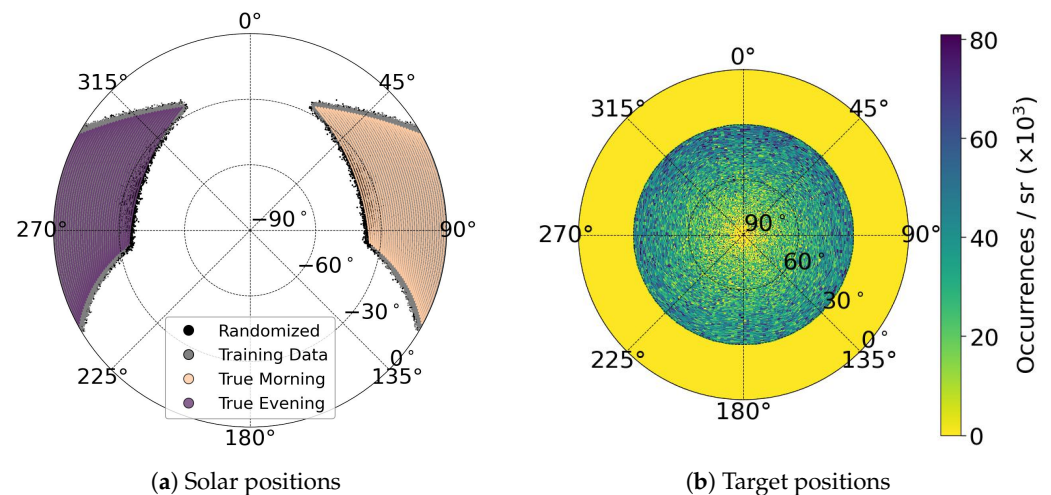


Figure 6. Solar azimuth and elevation angles (a) and target azimuth and elevation angles (b) for a specified ground site. Solar positions are displayed using a scatter plot, while target positions are displayed as a heatmap with purple indicating zero point density and red indicating maximum.

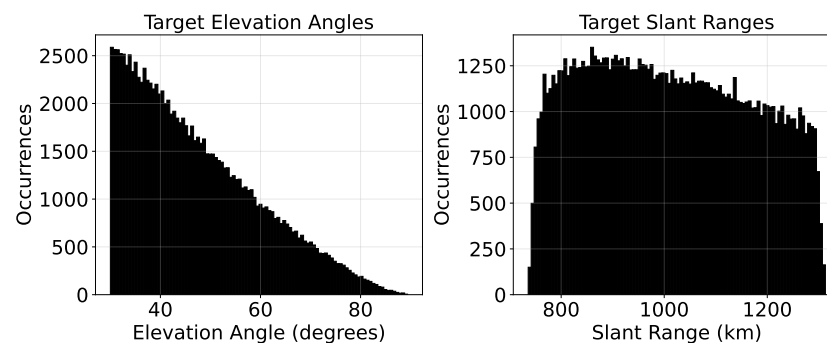


Figure 7. Histograms of training set target object elevation angles (left) and camera-target slant ranges (right). The distribution of elevation angles was intentionally biased to produce a more uniform distribution of slant ranges.

To stabilize optimization under megameter-scale Z and decimeter-scale X/Y , a SITE-style normalization is adopted from [5,6] and adapted for this work, using the slant range to calculate a target-plane HFOV in meters, normalize by the minimum HFOV, then divide by the RoI zoom ratio (Section 2.5.4). Coudé rotation is accounted for by (i) test-time derotation, (ii) uniform in-plane rotations applied to the training data illumination vectors, or (iii) telescope-specific analytic angles derived from az/el (preferable when the coudé path is known).

2.3. Synthetic Training Data Generation

With the imaging geometry established, this section details the pipeline used to generate the richly annotated synthetic dataset. The process begins with an accurate 3D CAD model and systematically applies realistic, application-specific degradations designed to bridge the Sim2Real domain gap.

2.3.1. Scene Configuration

Seasat CAD imagery is rendered for 110,000 unique 6DOF pose+illumination pairs at 256×256 px (80%/10%/10% train/validation/test). For each pose, EEVEE produces a partially illuminated render and a paired fully front-illuminated render (for relighting supervision). Material properties were inspired by [22] with randomized per-sample perturbations to reflect uncertain real-world material properties.

2.3.2. Sampling of Pose, Illumination, and Range

Rotations are sampled uniformly on $SO(3)$ ([23]), yielding the expected quaternion element histograms after normalization (Figure 8). World X/Y translations are approximately uniform in angle-space with the constraint of maintaining the complete object within the field of view (FOV), as displayed in Figure 9. In practice, sensor operators “click-track” if necessary to ensure objects are fully within the FOV. Target elevation is constrained to $\geq 30^\circ$ (image quality is significantly degraded $<30^\circ$ due to increased target range and atmospheric path length); target zenith angles (Figure 7) and Sun positions are as described previously.

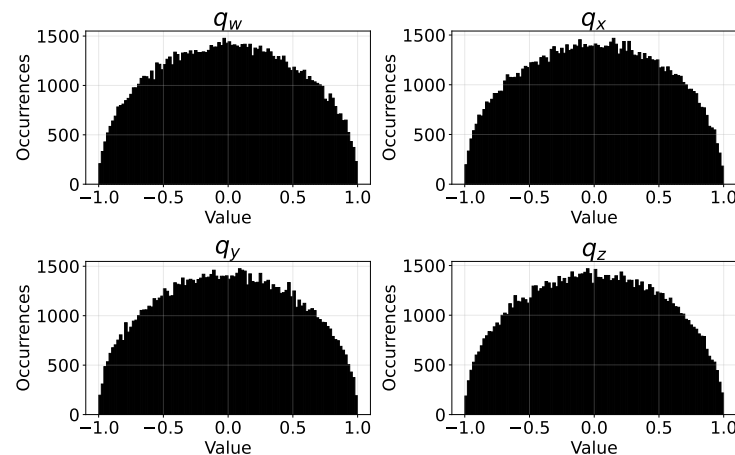


Figure 8. Histograms of training set quaternion elements (q_w , q_x , q_y , q_z). Quaternions were uniformly sampled over $SO(3)$ [23]. Note that uniform sampling in $SO(3)$ does not produce uniform distributions in quaternion element space.

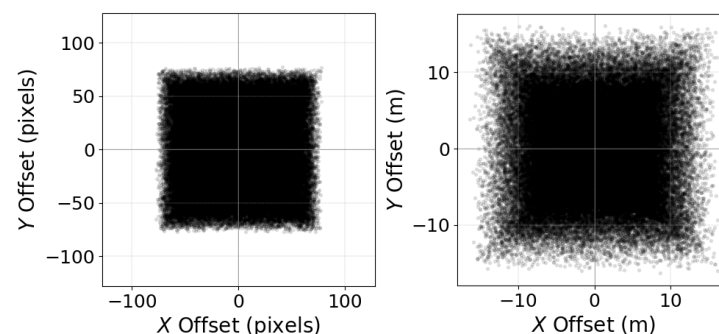


Figure 9. Image-plane translation offsets for HST training data shown in both pixel coordinates (left) and meters (right). The distributions are largely uniform until approaching the edges of the FOV where uniformity is reduced to meet the constraint of maintaining the full object within the FOV.

Slant range d is approximated with the spherical Earth form [24]:

$$d = \sqrt{R_E^2 \sin^2 \alpha + h^2 + 2hR_E - R_E \sin \alpha}, \quad R_E = 6371 \text{ km}, \quad (1)$$

with elevation angle α and satellite altitude h . In this specific application, the spherical Earth approximation results in slant range errors of less than 0.1%, which was deemed acceptable. Future work may incorporate a more accurate model such as WGS 84.

2.3.3. Rendered Supervision and Label Construction

For each sample the following are produced: RoI-cropped pristine image; RoI-cropped pristine front-lit image; dense per-pixel M_{XYZ} world coordinates; M_{SRA} 33-class surface-

region labels obtained via farthest-point sampling on the fused point cloud [5]; and p_{xy} maps that encode the full-frame coordinates of RoI pixels (Figures 10–12). These labels drive the staged training of the PoseNet (Section 2.5.3). The number of M_{SRA} regions was set to 32 (plus one background class) to balance accuracy and computational cost, a choice informed by the ablation study in [5] which demonstrated 70.5% accuracy versus 71.0% accuracy for 32 and 64 regions, respectively.

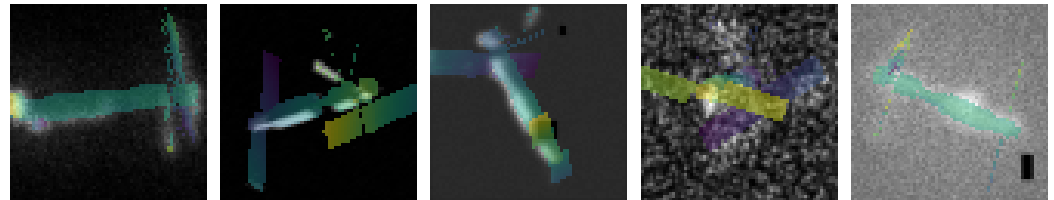


Figure 10. The ground truth Z components of M_{XYZ} 3D point maps are visualized using a colormap with partial transparency (yellow indicates pixels close to the camera, purple indicates pixels far from the camera), overlaid on their corresponding grayscale RoI-cropped synthetic images, showing dense pixelwise 2D↔3D correspondences. These data were generated for coupled pipeline training, where the RoI crop algorithm relied on localizations predicted by the Coarse Localizer. As a result, inaccuracies are evident in the crops.

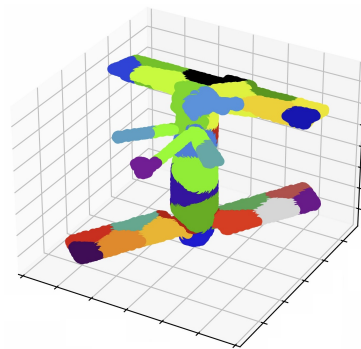


Figure 11. A 3D visualization of the Seasat point cloud (created by aligning and combining all M_{XYZ}) after application of farthest point sampling to segment the geometry and create M_{SRA} . The colors distinguish the different segmented surface regions.

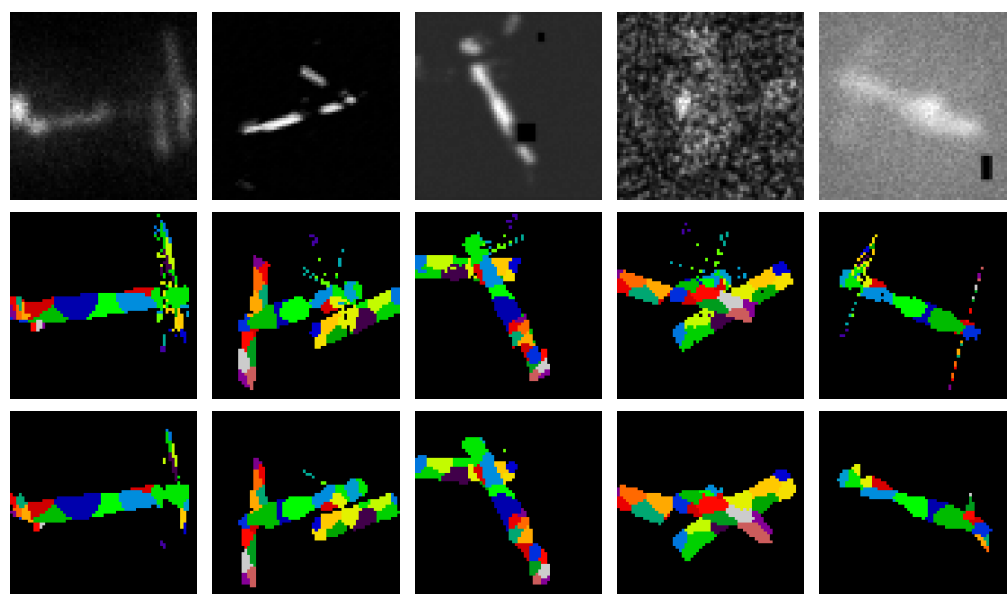


Figure 12. The first row shows the input RoI cropped test images. The second row displays the corresponding ground truth M_{SRA} , where each pixel is assigned one of 33 semantic classes (class 0 for background). The third row shows the corresponding ground truth M_{XYZ} , where each pixel is assigned one of 33 semantic classes (class 0 for background).

the background in black and 32 object-specific classes in distinct colors). These semantic classes provide a crucial intermediate representation, simplifying the downstream task of regressing the 6DOF pose by first requiring the network to solve the simpler problem of classifying which satellite region each pixel belongs to. The third row illustrates the maps predicted by the U-Net. The U-Net produces per-pixel 33-dimensional vectors via a softmax activation function in the final layer. Each element in the vector represents a softmax-normalized score for one of the 33 classes, indicating the model's relative confidence for that class. For visualization, these vectors were reduced to their maximum-valued class (the *argmax* over the 33 classes), and each class was assigned a unique color.

2.3.4. AO-Style Degradations and Augmentations

Inspired by [14], application-specific randomized augmentations are applied to the pristine renders to better simulate the appearance of real AO imagery. These augmentations are critical for bridging the Sim2Real domain gap, as underscored by SPEC2021 [13]. A 2D double-Gaussian PSF is used to approximate time-averaged AO blur. Ghost images are randomly synthesized by summing spatially shifted and scaled variants of the PSF to emulate high-frequency jitter. Random specular glints are also injected as isolated 2D Gaussian peaks positioned within the satellite bounding box, simulating transient solar reflections from satellite surfaces. A realistic sensor noise model combines four components: dark current, fixed pattern noise, Poisson-distributed shot noise, and Gaussian read noise. Additional noise augmentations include standalone Gaussian noise and multiplicative intensity noise (both pixelwise and image-wide). Brightness, contrast, gamma, and bias (pedestal) adjustments are also randomly applied. Other randomized degradations include additional motion blur, median blur, and Gaussian blur. Sensor artifacts such as hot pixels, dead pixels, and coarse dropout (random small rectangular occlusions) further increase realism and broaden the distribution of the training data (with the goal of encompassing the distribution of possible real imagery). Each transformation is applied randomly and independently with a tuned probability. The resulting training set includes 264,000 images, derived from 88,000 unique pose and illumination combinations. Low Fidelity In-Distribution (LFID) validation and test sets, each comprising 11,000 unique samples, are degraded using the same augmentation pipeline. Neural style transfer [14,16,25,26] was considered for future use but omitted here due to the limited availability of real AO satellite imagery. While applying augmentations dynamically at batch draw time can be more robust, a fixed set of augmentations was pre-generated for this work to simplify prototyping and coupled training between the Coarse Localizer and PoseNet. The double-Gaussian blur, ghost images, specular glints, sensor noise model, bias, hot pixel, and dead pixel implementations were custom while the remaining augmentations used [27]. All augmentation parameter ranges and probabilities were established prior to HFWO and real-image evaluation based on general knowledge of AO SDA imagery. No real evaluation imagery—nor imagery from the same satellite passes—was used to tune, fit, or inform the augmentation distributions, loss functions, or network hyperparameters, supporting a leakage-free Sim2Real evaluation. Furthermore, for the qualitative SPARS evaluation set, the primary model developer (first author) had no access to the 250 evaluation frames, which were independently scored by the second author.

2.3.5. Randomized MLI Texture

To account for the wrinkled multi-layer insulation (MLI) blankets characteristic of modern spacecraft (and visible in pre-launch images of HST), a bump-mapped noise field is procedurally randomized per render across seven Blender parameters (mapping X/Y scale, noise scale/roughness/distortion, bump strength/distance) as displayed in Figure 13 and detailed in [18]. The real HST test images were not used to inform MLI randomization.

These parameters were randomized across broad bounds rather than fit to specific real image sets.

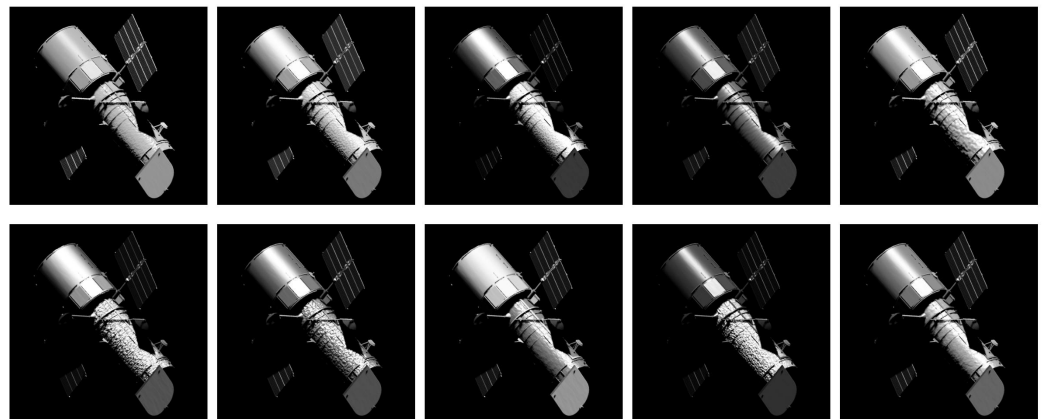


Figure 13. Ten HST renders showing randomized material properties, illumination strength and diffuse/specular balance, and multi-layer insulation (MLI) surface texture. Camera parameters, object pose, and illumination direction are held constant for display purposes.

2.4. Synthetic Test Data Generation

Due to the limited availability of labeled real imagery for model validation, high-fidelity synthetic test data were created using a fully independent process to create a simulated domain gap (Sim2Sim). Digital Imaging and Remote Sensing Image Generation (DIRSIG) was used to render object planes and High Contrast Imaging for Python (HCIPy) was used to generate realistic AO PSFs and realistic noise via a simulated closed-loop AO system using wave optics and random atmosphere with temporal coherence [28–30]. A notional AO system was simulated with specifications similar to [31,32]. The HCIPy simulation produced broadband PSFs with enhanced realism and complexity compared to the empirical Gaussian PSFs used for the training data. This pipeline controlled for atmospheric conditions such as the Fried parameter, r_0 [33,34], enabling quantification of relationships between pose model accuracy and atmospheric turbulence and image quality. Noteworthy enhancements for the DIRSIG renders are:

- Inclusion of Earth’s surface as an additional illumination source.
- Use of a physics-based MODTRAN atmosphere model.
- Lunar contributions to both direct object illumination and atmospheric path radiance.
- Support for multi-bounce surface reflections.
- Physics-based camera model with perspective projection.
- Extended Sun source (0.5°) producing soft shadow edges.
- Validated computation of 32-bit at-aperture radiance in $\text{W}/\text{cm}^2/\text{sr}/\mu\text{m}$.

In contrast, the Blender simulation lacked Earth and lunar illumination sources, did not simulate atmospheric effects or path radiance, excluded multi-bounce reflections, used a simple scaled orthographic projection, treated the Sun as a point source, and output unitless 8-bit pixel values. A comparison of Seasat renders produced with Blender and DIRSIG is shown in Figure 14. While effective at synthesizing labeled, highly realistic SDA AO imagery for model validation the Blender+DIRSIG pipeline currently requires >10 s of computation per frame, making it impractical for synthesizing viably sized training sets. An HCIPy-simulated AO PSF is displayed in Figure 15. HFWO images are displayed in Figures 16–19.

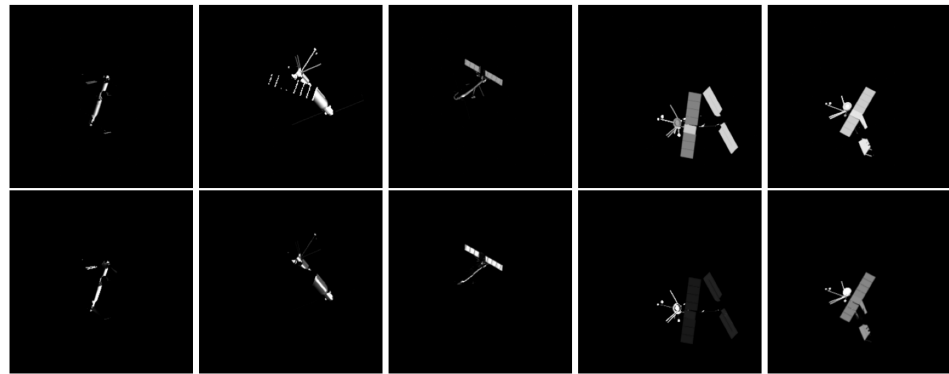


Figure 14. Five pristine test image renders produced with Blender (top five) and Digital Imaging and Remote Sensing Image Generation (DIRSIG) (bottom five). Despite identical CAD geometry, satellite pose, and illumination directions, the renders appear noticeably different due to DIRSIG's more advanced simulation pipeline and intentionally distinct material properties designed to stress-test the model's generalization capability. Renders are displayed with an individual linear min-max stretch.

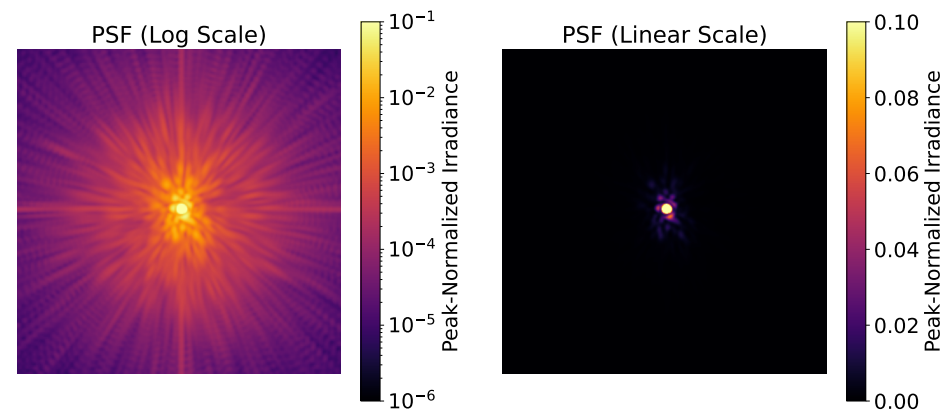


Figure 15. A point spread function (PSF) simulated using High Contrast Imaging for Python (HCIPy), displayed in $\log_{10}$ scale at (left) and linear scale at (right). Complex structure is visible in the PSF. This method is capable of producing highly complex and realistic PSFs.

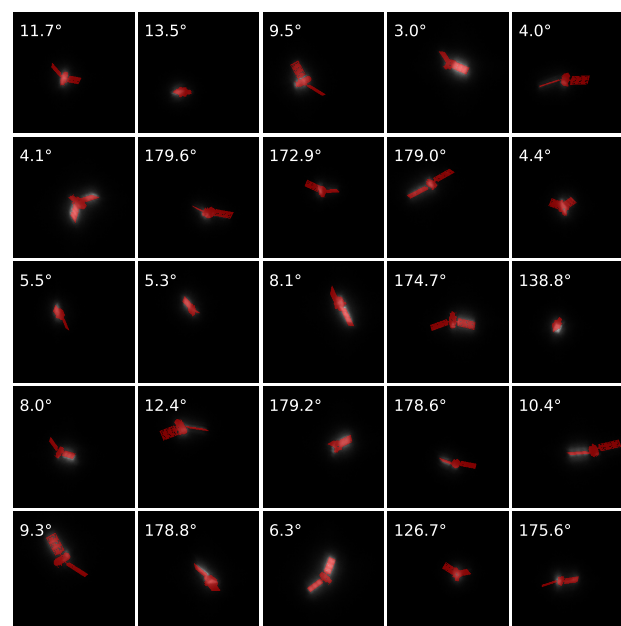


Figure 16. Twenty-five HFWO ARGOS test images with $r_0 = 6$ cm displayed in grayscale, and renders produced with the pose model estimates overlaid as red wireframes. Rotation errors are shown overlaid at the upper-left of each image.

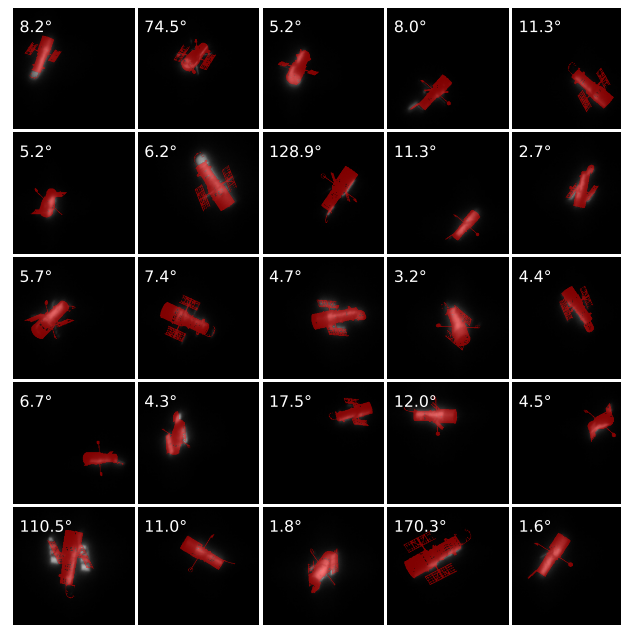


Figure 17. Twenty-five HFWO HST test images with $r_0 = 6$ cm displayed in grayscale, and renders produced with the pose model estimates overlaid as red wireframes. Rotation errors are shown overlaid at the upper-left of each image.

Maximum vs. Minimum Error Class Examples

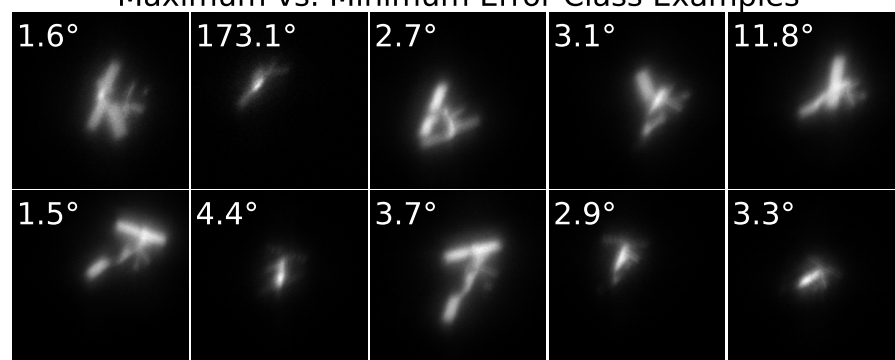


Figure 18. The first row displays five test images from the pose class with the maximum mean rotation error, with per-sample rotation errors overlaid at the upper left of each test image. The second row displays five test images from the pose class with the minimum mean rotation error.

Maximum vs. Minimum Error Class Examples

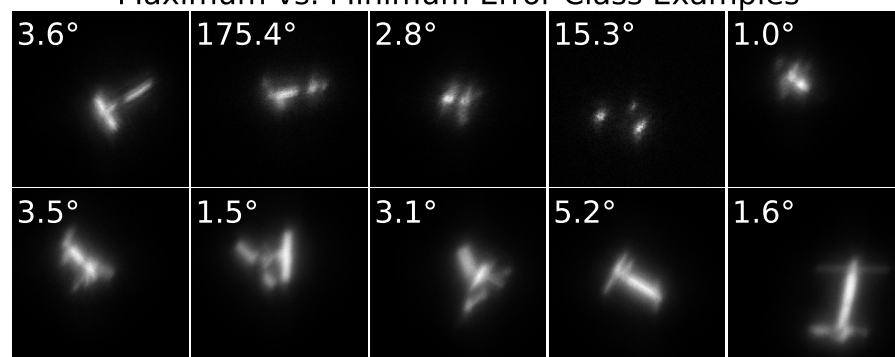


Figure 19. The first row displays five test images from the illumination direction class with the maximum mean rotation error, with per-sample rotation errors overlaid at the upper left of each test image. The second row displays five test images from the illumination direction class with the minimum mean rotation error.

2.5. Learning System

2.5.1. Coarse Localizer

The localizer (Figure 4) regresses amodal boxes ($x_{\min}, y_{\min}, x_{\max}, y_{\max}$) and object center (x, y) in continuous image coordinates: $[0, 1]$. Inputs are the 256×256 grayscale image, a 3-channel illumination vector (constant across the grid, normalized to $[0, 1]$), and an orthographic scale estimate similarly replicated across the full image-dimension grid to fill a channel. A U-Net backbone (~ 8 M parameters) first predicts a reconstructed + relit image; two heads (~ 8.5 M parameters each) then regress the box (Complete Intersection over Union (CIoU) loss) and center (Mean Squared Error (MSE) loss) [35,36]. Amodal supervision enables the RoI crop to center the CAD origin, reducing residual translation for PoseNet. A custom Coarse Localizer was used in contrast to [5] due to the unique properties of this imagery, the fact that Faster R-CNN [37] required additional training time, and to take advantage of additional application-specific inputs (target range, illumination direction).

2.5.2. RoI Crop, Resize, and p_{xy} Maps

The RoI algorithm enforces: (1) square crop before resize; (2) RoI contains the full predicted amodal bounding box; (3) predicted center at RoI center. Zero-padding is applied if needed. The crop is resized to 64×64 and paired with p_{xy} maps ($64 \times 64 \times 2$) that preserve each pixel's location in the original frame, serving as dense 2D keypoints for $2D \leftrightarrow 3D$ correspondences. Training is coupled: PoseNet consumes localizer predictions rather than randomly perturbed ground truth, matching test-time failure modes (Figure 20). This is done because of the broad distribution of image quality produced by AO instruments and the strong correlation between localizer performance and image quality. Coupled training preserves this correlation, though special care must be taken in training not to overfit the model to the training set.

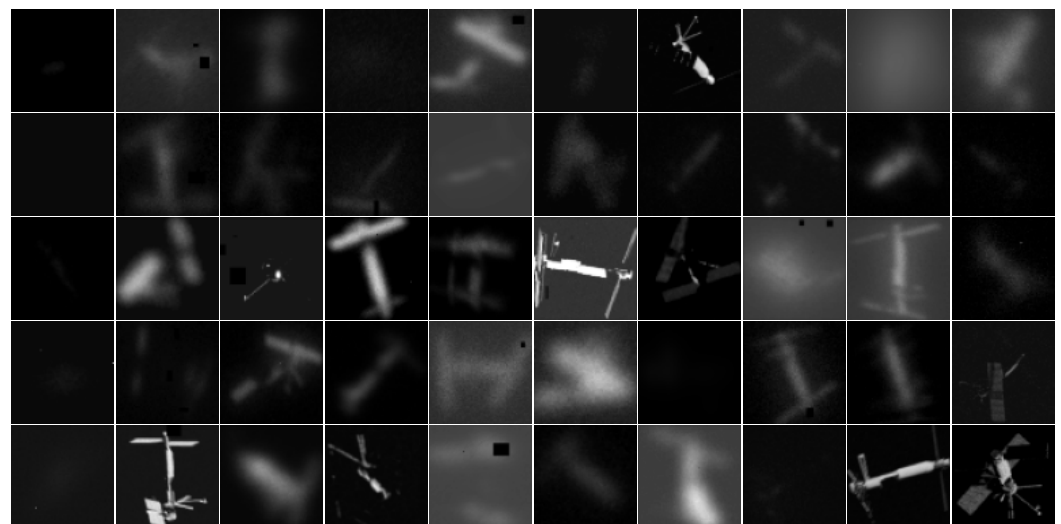


Figure 20. The first 50 Low Fidelity In-Distribution (LFID) test images from the augmented 6DOF Seasat dataset after the RoI crop and resize algorithm has been applied using the predicted bounding box and object center coordinates.

2.5.3. PoseNet for 6DOF (Extendable to n -DOF)

PoseNet (Figure 5) comprises four compact U-Nets (backbone) feeding a convolutional neural network (CNN) pose regressor head:

1. Reconstruction U-Net: inputs = RoI + 3-ch illum. vector; output = denoised/deblurred image.

2. Relighting U-Net: inputs = RoI + 3-ch illum. vector + denoised/deblurred image; output = fully front-lit (deshadowed) image.
3. M_{SRA} U-Net (32 object surface region classes, 1 background class): inputs = RoI + reconstruction output + relighting output; output = class logits.
4. M_{XYZ} U-Net: inputs = M_{SRA} + RoI + p_{xy} ; output = dense world X/Y/Z (set-normalized to range $[0, 1]$).

The regressor ingests a 41-channel tensor (concatenating RoI, reconstruction output, relighting output, M_{SRA} , M_{XYZ} , p_{xy}). A 7D auxiliary vector [scale, $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$, (x, y)] is provided as input at the feature vector (first dense layer). Orientation uses the continuous 6D representation [11] followed by Gram-Schmidt orthogonalization to $SO(3)$; translation t_{XYZ} is linear. The full model has ~ 71 M parameters; further increasing parameters by adding nodes to the dense layers while holding the training set constant did not improve model accuracy.

2.5.4. Scale Normalization

A SITE-style scheme [5,6] is adapted to orthography and kilometer-scale Z: compute the range-derived target-plane HFOV, normalize by the minimum HFOV, then divide by the RoI zoom ratio. This keeps the Z channel values in family with X/Y offsets for stable training.

2.5.5. Training Schedule and Losses

Optimization uses Adam [38] with the default initialized learning rate of 1×10^{-3} unless otherwise stated and all training employs early stopping (patience = 10) on validation loss. Stage 1: train reconstruction+relighting U-Nets jointly with a combined loss; freeze. Stage 2: train M_{SRA} ; freeze. Stage 3: train M_{XYZ} , briefly unfreeze upstream for end-to-end finetuning with learning rate initialized to 1×10^{-4} ; refreeze. Stage 4: train the CNN pose regression head (Figure 21). Losses are: MSE (reconstruction/relighting), sparse categorical cross-entropy (M_{SRA}), Mean Absolute Error (MAE) (M_{XYZ}). The final disentangled pose loss is

$$\mathcal{L}_{6DOF} = \mathcal{L}_R + \mathcal{L}_{\text{center}} + \mathcal{L}_z, \quad (2)$$

without weighting, as adopted from [5].

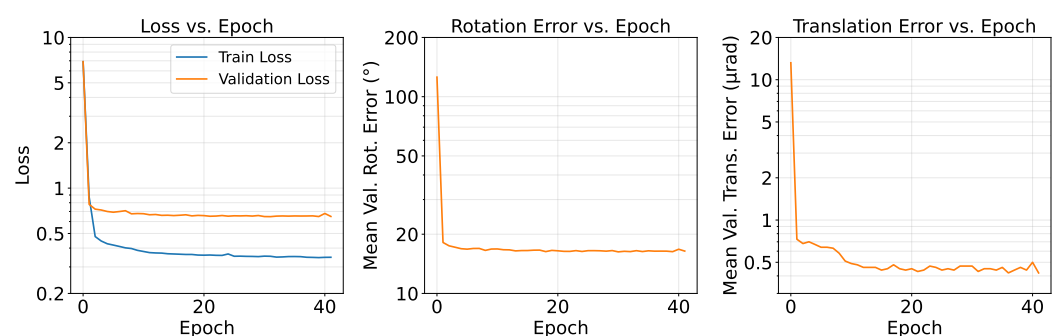


Figure 21. Training and validation loss as a function of epoch (left), mean LFID validation rotation error in degrees (center), and mean LFID validation two-axis angular translation error in μrad (right) during Stage 4 of PoseNet training. Validation metrics improve rapidly during the first five epochs. Weights were checkpointed based on the best validation loss, and training terminated early (prior to 50 epochs) using a patience of 10 epochs without validation loss improvement.

2.5.6. Articulations (7DOF)

To incorporate a single additional rotational degree of freedom, the disentangled pose loss in Section 2.5.5 (Equation (2)) is extended with a symmetry-aware term using a continuous sine-cosine representation $x = (\sin \theta, \cos \theta)$:

$$\mathcal{L}_{7\text{DOF}} = \mathcal{L}_R + \mathcal{L}_{\text{center}} + \mathcal{L}_z + \mathcal{L}_x, \quad \mathcal{L}_x = \min(\|\hat{x} - x\|_1, \|\hat{x} + x\|_1), \quad (3)$$

where $\hat{x} = (\sin \theta_{\text{pred}}, \cos \theta_{\text{pred}})$ and $x = (\sin \theta_{\text{true}}, \cos \theta_{\text{true}})$. The network outputs continuous predictions via a two-node tanh activation head. At inference, predicted angles are extracted from the continuous output vector $\hat{x}$ using $\hat{\theta} = \text{atan2}(\sin \theta_{\text{pred}}, \cos \theta_{\text{pred}})$.

The 180° antipodal symmetry of the solar array is explicitly handled during training via the $\min(\cdot)$ operator in Equation (3), which computes the L_1 norm between predictions and both the true and sign-flipped target vectors. This formulation forces loss invariance between θ and $\theta \pm 180^\circ$. Directly training the network without this invariant loss failed to converge due to conflicting gradients at antipodal orientations. For evaluation and error reporting, this 180° ambiguity is similarly accounted for using a symmetry-adjusted metric $\theta_e = \min(\theta_{\text{err}}, 180^\circ - \theta_{\text{err}})$, where $\theta_{\text{err}} = |\hat{\theta} - \theta_{\text{true}}|$.

On pristine synthetic 7DOF Seasat test imagery, the raw solar array articulation error is 90° (equivalent to uniform guessing), whereas the symmetry-adjusted error drops to 26.4°, confirming successful convergence on symmetry-invariant rotation. Architecturally, the solar array regression branch ingests both an intermediate feature vector and the predicted main-body orientation (displayed in Figure 22), providing body-frame reference awareness. The framework may be extensible to n -DOF by adding parallel regression branches (e.g., concatenating primary translation predictions into a single linear head for 1DOF linear components), though performance will degrade as n increases. While the pipeline does not resolve absolute, symmetry-broken orientation end-to-end, the resulting pose estimates remain operationally useful when antipodal orientation equivalence is acceptable or when external priors (such as illumination direction at inference time) are applied to break symmetry.

2.6. Kalman Filtering and Range Post-Processing

Image sequences pass through the Coarse Localizer, RoI crop, PoseNet, and range post-processing pipeline. End-to-end pose estimation executes at 7.1 Hz; Kalman filtering adds negligible computational overhead (~10 ms per frame), maintaining equivalent throughput.

2.6.1. Filtering Architecture and Continuous State Representations

To enforce temporal consistency across sequential images while maintaining causal operation, predictions are filtered in two stages:

1. **Coarse Localizer Output Filtering:** The bounding box upper-left/lower-right coordinates $\mathbf{b} \in \mathbb{R}^4$ and object center coordinates $\mathbf{c} \in \mathbb{R}^2$ are filtered in normalized image space $[0, 1]$ prior to RoI cropping. To mitigate image-plane spatial inference jitter without introducing spurious cross-axis dependencies, $\mathbf{b}$ and $\mathbf{c}$ are each filtered independently using diagonal covariance matrices. Component coupling occurs exclusively through a joint statistical outlier gate.
2. **PoseNet Output Filtering:** Filtering is selectively applied to the highly non-linear rotational and articulation states. Translational states output from the PoseNet (X, Y, S) are intentionally passed through unfiltered to prevent phase lag.

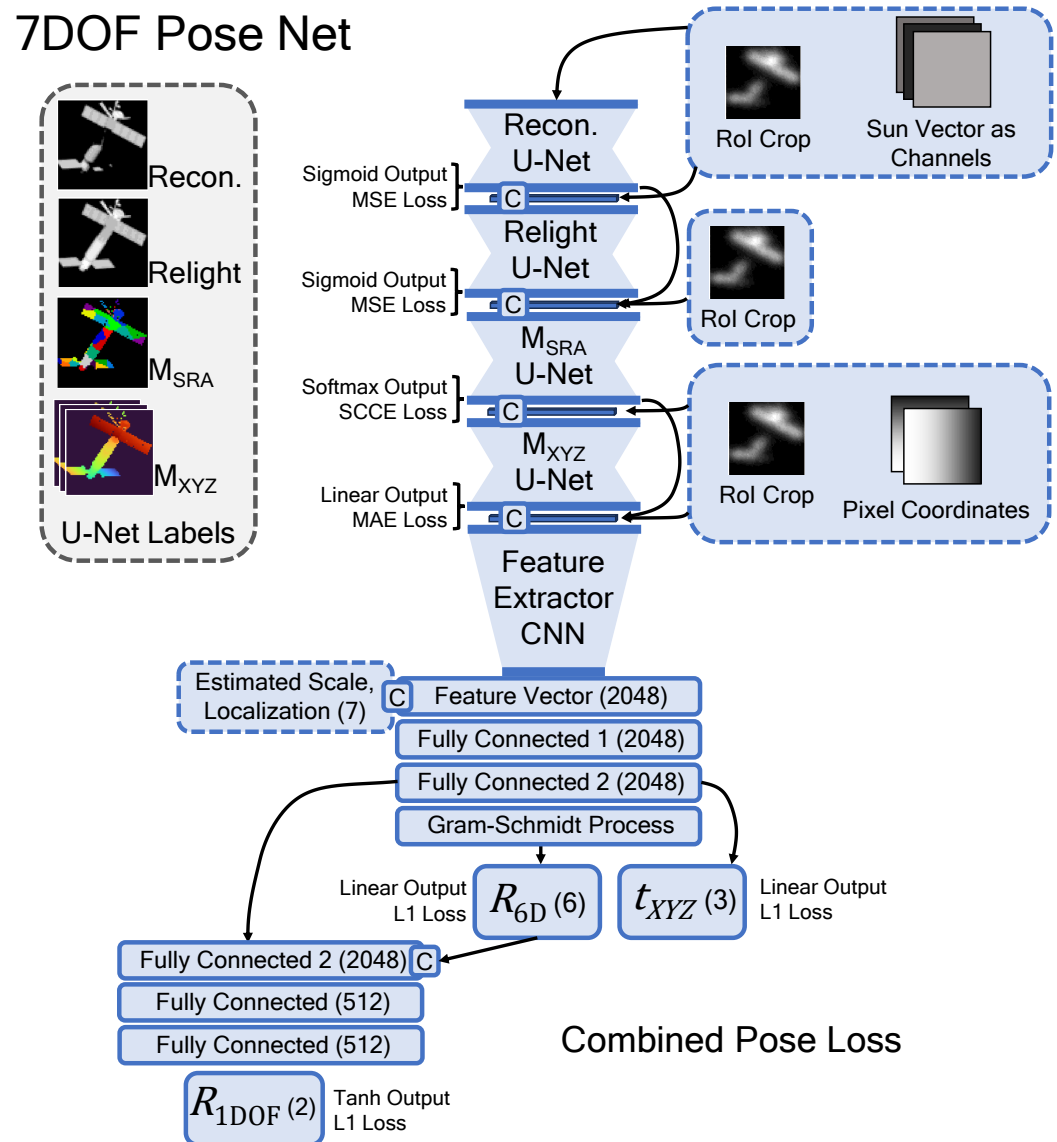


Figure 22. Diagram showing the architecture of the 7DOF network, extendable to n -DOF with additional branches. For comparison, the 6DOF architecture is displayed in Figure 5.

Filtering standard 3DOF rotation representations introduces gimbal lock, antipodal sign ambiguities ($q \equiv -q$), and angular wrap discontinuities. Therefore, PoseNet output filtering operates on carefully selected continuous representations:

- **3DOF Rotation:** Filtered in the continuous 6D rotation representation $\mathbf{r} = [v_1, v_2] \in \mathbb{R}^6$ [11]. The full 6D vector is jointly filtered using full cross-covariance matrices to account for multivariate correlations. Gram-Schmidt orthogonalization onto $SO(3)$ and quaternion conversion are performed after filtering.
- **1DOF Solar Array Articulation:** As described in Section 2.5.6, PoseNet is deliberately trained to be invariant to the solar array's 180° antipodal symmetry (Equation (3)), so consecutive raw $[\sin \theta, \cos \theta]$ predictions can legitimately flip by π even under continuous physical motion. These flips are not measurement noise and must not be smoothed away by the filter. Articulation is therefore filtered in the doubled-angle representation $[\sin 2\theta, \cos 2\theta] \in \mathbb{R}^2$, removing the ambiguity from the filter's input rather than attempting to detect and reject it after the fact. These components are jointly filtered as a 2D multivariate state using full cross-covariance. The filtered estimate

is then reconverted, recovering one representative of the $\theta, \theta + \pi$ equivalence class, consistent with the symmetry-adjusted metric θ_e used for evaluation.

2.6.2. Uncertainty-Aware Kalman Filter Formulation and Outlier Rejection

Temporal filtering employs a discrete-time linear Kalman filter [39] using a zero-order (Random Walk) state-transition model,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{w}_k, \quad \mathbf{z}_k = \mathbf{x}_k + \mathbf{v}_k, \quad (4)$$

where $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ and $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$. The zero-order model is deliberately selected to prevent non-linear extrapolation errors on the constrained $SO(3)$ rotational manifold.

To accommodate inference noise, the process ($\mathbf{Q}$) and measurement ($\mathbf{R}$) covariance matrices are adaptive. $\mathbf{Q}$ is computed using the sample covariance of the frame-to-frame differences of the raw, pre-filter measurement sequence ($\Delta \mathbf{z}$), serving as a proxy for the kinematic velocity, and $\mathbf{R}$ is computed using the sample covariance of the measurement residuals ($\mathbf{z} - \hat{\mathbf{x}}$), both evaluated over a causal sliding window of $W = 100$ frames. Furthermore, for PoseNet outputs, the measurement noise covariance is dynamically inflated frame-by-frame as $\mathbf{R}_k = \mathbf{R}/c_k$, where $c_k \in [0.01, 1.0]$ is a scalar confidence coefficient derived from the spatial-average Shannon entropy of the M_{SRA} softmax distribution (an intermediate output of PoseNet inference). Per-pixel entropy $H_{i,j} = -\sum_{m=1}^{33} p_{i,j,m} \ln p_{i,j,m}$ is averaged across the RoI and normalized against the maximum theoretical entropy ($\ln 33$) to obtain a raw confidence value; values below a raw confidence of 0.65 are floored, and the remaining range is linearly rescaled to yield the final normalized confidence score c_k . As visual ambiguity increases spatial entropy ($c_k \rightarrow 0$), $\mathbf{R}_k$ grows, with the intention that the filter discounts uncertain predictions. Minimum covariance floors of $\mathbf{Q}_0 = \mathbf{R}_0 = 10^{-2}\mathbf{I}$ are strictly enforced for all filters.

Two sequential gates isolate and reject significant inference noise:

1. **Physical Rate Limits:** The raw innovation magnitude is strictly bounded prior to the statistical update step. For the 6D representation for 3DOF rotation, this acts as a maximum inter-frame geodesic rotation angle on $SO(3)$ (equivalent to the quaternion distance) configured to 10° , reflective of typical operational framerates and reasonable maximum apparent satellite rotation rate. Similarly, the solar array articulation innovation magnitude is bounded based on a 1° maximum inter-frame rotation, based on reasonable maximum solar array drive speeds.
2. **Statistical Outlier Gating:** A Normalized Innovation Squared (NIS) gate applies a chi-square (χ^2) hypothesis test ($\alpha = 0.90$) [40]. Measurements exceeding this threshold are soft-clipped to lie exactly on the statistical boundary, ensuring bounded innovation while aiming to discount erroneous network predictions.

All results reported in this work employ strictly forward-causal filtering operating in real time without access to future frames. To prevent data leakage, filter hyperparameters (sliding-window length, covariance floors, and NIS confidence thresholds) were established as generalized physical and statistical baselines and fixed. No filter hyperparameters were tuned on real test data. This uncertainty-aware filtering pipeline attempts to provide increased operational robustness on real-world imagery, with results presented in Section 3.1.3. It remains a future-work priority to create a set of synthetic time-series data sufficient for filter tuning and further filter development.

2.6.3. Range Post-Processing

Range post-processing blends network-predicted scale with ephemeris priors according to the combined operational uncertainty of TLE, IFOV, and CAD model fidelity. Predictions are handled in three regimes based on prior uncertainty:

1. High Uncertainty: Use the raw predicted scale directly.
2. Moderate Uncertainty: Clip the predicted scale to within $\pm 2.5\%$ of the input estimate, matching the bounds used during training.
3. Low Uncertainty: Retain the ephemeris prior directly.

All results reported in this work use the moderate-uncertainty implementation. In practical operations, CAD model dimensional mismatch is expected to dominate range error. This systematic bias can be compensated either dynamically by scaling the target CAD dimensions using the ratio of the TLE-derived range to the predicted range, or permanently by refining the CAD geometry once systematic discrepancies have been quantified. The disparity between lateral (X/Y) and target slant range (Z) errors is inherent to monocular ground-based AO imaging: a ~ 1 px image-space error corresponds to approximately 10 cm lateral error, but approximately 10 km slant range error at $Z \sim 1000$ km for a 100 nrad IFOV [18].

2.7. Evaluation Datasets and Metrics

Three sources are used: (1) LFID synthetic test set (11,000 unique samples) generated with the same process as for training; (2) HFWO synthetic test set hold-outs produced with DIRSIG+HCIPy as described; and (3) real AO imagery. Metrics follow [5]: rotation via quaternion distance (deg); image-plane translation as two-axis angular error (nrad) or as pixel error on the 256×256 frame or in meters; and range in kilometers or as relative error (%). Unless otherwise stated, evaluations use the coupled pipeline (Section 2.5.2); decoupled ablations are displayed in Figure 23.

To quantify rotation error, quaternion distance (also called rotation magnitude or geodesic error [11,41]) is used [12,13]. The quaternion distance e_q between a ground-truth rotation q and an estimated rotation $\hat{q}$, both unit quaternions, is computed as:

$$e_q = 2 \arccos(|\langle \hat{q}, q \rangle|). \quad (5)$$

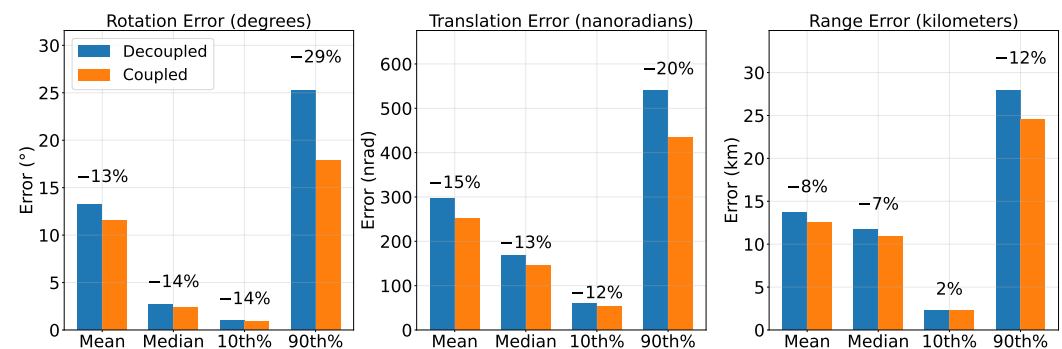


Figure 23. Comparison of decoupled (blue) and coupled (orange) model performance. The percentage change from decoupled to coupled is shown in black above each of the 12 metrics. The greatest impact was to the 90th percentile rotation error, with a 29% decrease. Performance improved for 11 of 12 metrics, with a 2% increase in 10th percentile range error.

2.8. Key Approach Adaptations

Architecturally, the 6DOF PoseNet (Figure 5) introduced several key adaptations beyond GDR-Net [5]. The model incorporates a TLE-derived range estimate and illumination vector, improving performance. Normalization was adapted for ~ 1000 km target ranges. An auxiliary localization vector was injected at the pose head to refresh memory, reducing reliance on high spatial frequency features surviving backbone propagation. Image reconstruction and relighting subnetworks were added to better handle degraded imagery. A serial backbone of four smaller U-Nets replaced the single large ResNet-34

backbone in [5], improving stability under degraded conditions by progressing from simpler tasks (denoising, deblurring) to more complex ones (deshadowing, classification, regression), and enabling efficient convergence on a single NVIDIA A100 GPU. A coupled training pipeline passed predictions from a custom Coarse Localizer (Figure 4) directly into the pose model, preserving correlations between localizer performance and image quality. To handle articulating satellites, a branched regression head and symmetry-aware loss extended the GDR-Net framework beyond 6DOF without necessitating additional model/object instances.

2.9. Implementation

Tooling includes Blender/EEVEE (v3.6.0), Astropy (v5.1), SciPy (v1.10.0), Albumentations (v1.3.1), TensorFlow/Keras (v2.6.0), and HCIPy (v0.5.1) which are all well-documented and publicly available. Training was conducted on a single NVIDIA A100 GPU. The HCIPy implementation closely follows [29]. The specific CAD models used in this work are property of a third party and cannot be distributed, though equivalent-quality CAD models of HST are freely available [42]. As described, the model architecture is adapted from GDR-Net [5] which is available [7,43], and the modifications are described in detail here and in [18]. The authors acknowledge Research Computing at the Rochester Institute of Technology for providing computational resources and support that have contributed to the research results reported in this publication [44].

3. Results

3.1. Real Imagery Results

3.1.1. Seasat (Framewise Inference)

The pipeline was evaluated on 137 manually-labeled real Seasat images (see Chapter 6 of [18] for additional details on real data and labeling), processed fully independently without Kalman filtering. Ground-truth pose labels were generated by manually aligning a 3D CAD model wireframe over target imagery in Blender. For video sequences, key frames were selected at regular intervals to capture major pose changes, with intermediate frames generated via spherical linear interpolation (slerp) for rotation and linear interpolation for translation. Manual alignment was performed without prior knowledge of sensor IFOV, target range, or illumination direction, introducing qualitatively estimated label noise of 1° – 5° . A similar procedure was applied to generate 7DOF ground-truth labels (including solar array rotation) for the 249-frame HST pass and single published frame.

Quantitative evaluation on these real collections is restricted to 5DOF spatial pose (3DOF attitude and 2DOF image-plane translation) for Seasat, and 6DOF (3DOF attitude, 2DOF image-plane translation, and 1DOF solar array articulation) for HST. Because ground-truth slant range and sensor IFOV were unreported for these published real datasets captured by Starfire Optical Range (SOR) and Advanced Electro-Optical System (AEOS), manual annotation of absolute range is ill-posed; a single-pixel error in target scale corresponds to multi-kilometer slant range variations at megameter distances (~ 1000 km). Consequently, during real-data quantitative testing, manually labeled orthographic scale values were supplied to the model as input proxies in place of TLE-derived slant range. These manual scale proxies introduced inherent label noise, which likely degraded the model's rotation and image-plane translation performance compared to operational deployment where accurate TLE-computed priors are available.

Table 2 summarizes the real datasets, and Table 3 details the performance metrics for Seasat, while qualitative pose prediction overlays are illustrated in Figure 24. Across the entire dataset, the model achieved a mean rotation error of 5.0° and a mean two-axis image-plane translation error of 21 cm. These numerical results incorporate the legacy

annotation noise discussed above, with the visual overlays serving as a qualitative baseline of model performance under real-world atmospheric degradations.

Table 2. Summary of real-image evaluation sets across satellite targets and sources.

Target	Dataset/Source	Evaluation Type	Images
Seasat	SOR 105-frame video [45]	Quantitative (5DOF) ^a	105
Seasat	SOR 30-frame video [46]	Quantitative (5DOF) ^a	30
Seasat	AEOS single frames [47,48]	Quantitative (5DOF) ^a	2
Subtotal Seasat Quantitative			137
HST	AEOS 249-frame pass (221 at $\geq 30^\circ$)	Quantitative (6DOF) ^b	249
HST	AEOS single frame [49]	Quantitative (6DOF) ^b	1
Subtotal HST Quantitative			250
Seasat	Multi-year SOR archive ($\geq 30^\circ$)	Qualitative SPARS	199
Total Real Test Images Evaluated			586

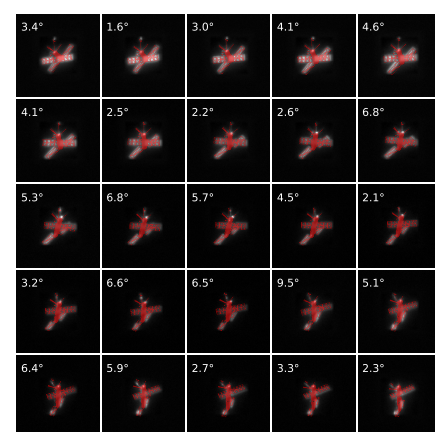
^a Evaluated dimensions are limited to 3DOF orientation and 2DOF image-plane translation. Slant range error was not quantified. ^b Evaluated dimensions are limited to 3DOF orientation, 2DOF image-plane translation, and 1DOF solar array articulation. Slant range error was not quantified.

Table 3. Rotation and translation error statistics for all 137 real Seasat test images, separated by dataset. The Werth et al. [47] and Fulcoly et al. [48] datasets are individual frames. For the results below all test frames were processed individually, with no Kalman filtering.

Dataset	Rotation Error ($^\circ$)				Translation Error (cm)			
	Mean	Median	Min	Max	Mean	Median	Min	Max
105-Frame Video	5.2	5.1	1.8	9.2	16	15	2.5	35
30-Frame Video	4.1	3.3	1.6	9.5	36	26	0.8	93
Werth et al.	2.1	–	–	–	59	–	–	–
Fulcoly et al.	8.2	–	–	–	39	–	–	–
All Combined	5.0	4.9	1.6	9.5	21	19	0.8	93
<i>Standard Deviation</i>	1.9	–	–	–	17	–	–	–



(a) 105-frame sequence.



(b) 30-frame sequence.

Figure 24. *Cont.*

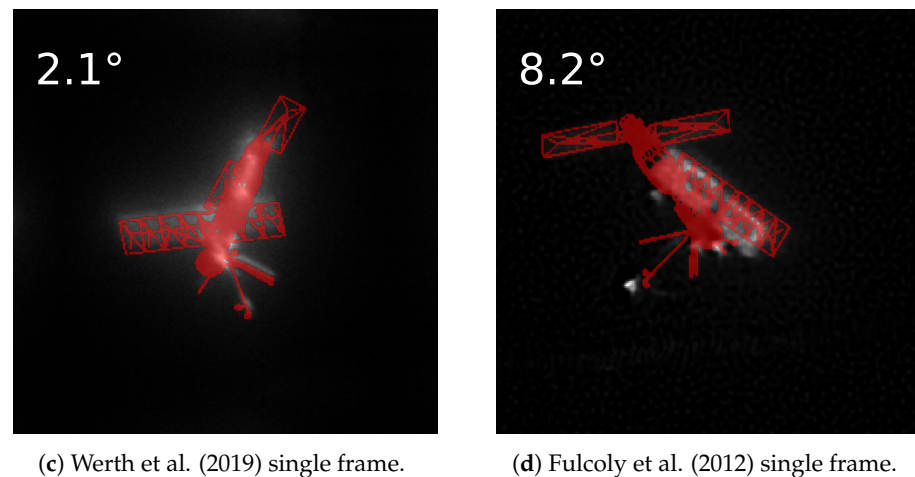


Figure 24. Pose predictions (red wireframes) overlaid on real Seasat images from SDA AO systems; approximate rotation errors shown at upper left of each panel. The images in (a), (c), and (d) were from [45], [47], and [48], respectively.

3.1.2. Qualitative Evaluation on Seasat (Frame-wise Inference)

Air Force Research Laboratory (AFRL) evaluated 250 additional real Seasat images (randomly sampled from >10,000 collected over nine years at one site). Frames were processed independently with a prior PoseNet architecture variant (lacking reconstruction/relighting U-Nets and illumination vector input). To avoid the challenge of accurate manual labeling for this additional data, we introduce the Satellite Pose Alignment Rating Scale (SPARS). SPARS is a five-level qualitative metric defined with visual examples in Figure 25; the resulting score distributions are shown in Figure 26 and Table 4.

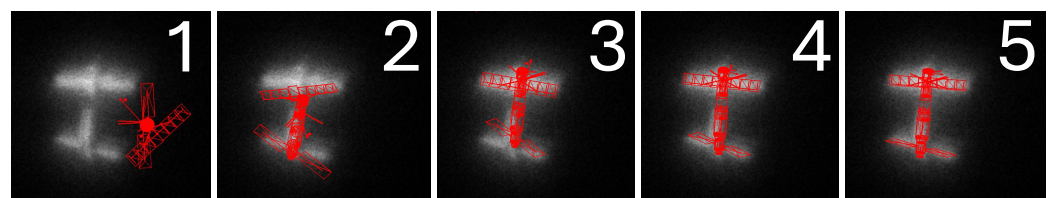


Figure 25. Red wireframe overlays on a grayscale test image showing notional pose estimates, evaluated left to right using the Satellite Pose Alignment Rating Scale (SPARS): (SPARS-1) catastrophic failure, no silhouette overlap; (SPARS-2) limited success, partial overlap but major misalignment; (SPARS-3) moderate success, good overlap with minor but noticeable errors; (SPARS-4) high-confidence match, excellent alignment with only subtle discrepancies; and (SPARS-5) ground truth equivalent, indistinguishable from true pose, even to a trained analyst. The SPARS category is annotated at the upper right of each panel.

The SPARS evaluations were performed frame-wise on isolated, non-sequential real imagery by a single domain-expert evaluator (the second author). To ensure a leakage-free evaluation, the primary model developer (first author) had no prior access to these 250 real evaluation images when configuring network architectures, loss functions, or synthetic training data augmentations. For each evaluated frame, the rater inspected the real AO image with the model's predicted 3D pose wireframe directly overlaid, assessing the visual fidelity of the alignment without access to underlying ground-truth reference data. Because this initial evaluation relied on a single rater, intra-rater repeatability and inter-rater agreement were not formally quantified. Establishing formal repeatability and reliability metrics across multiple independent analysts is an essential direction for future work to validate the standardization and operational utility of the SPARS qualitative scale.

The mean SPARS score across all 250 images was 3.90; below 30° elevation (unseen during training) the mean was 2.90 ($N = 51$), versus 4.16 at $\geq 30^\circ$ ($N = 199$). Aggregated, 78.4% achieved high-confidence or ground-truth-equivalent ratings (SPARS 4–5), with only 14% falling into SPARS 1–2 (71% of those at $<30^\circ$). These outcomes align with the framewise results above and reinforce robustness to real operational variability.

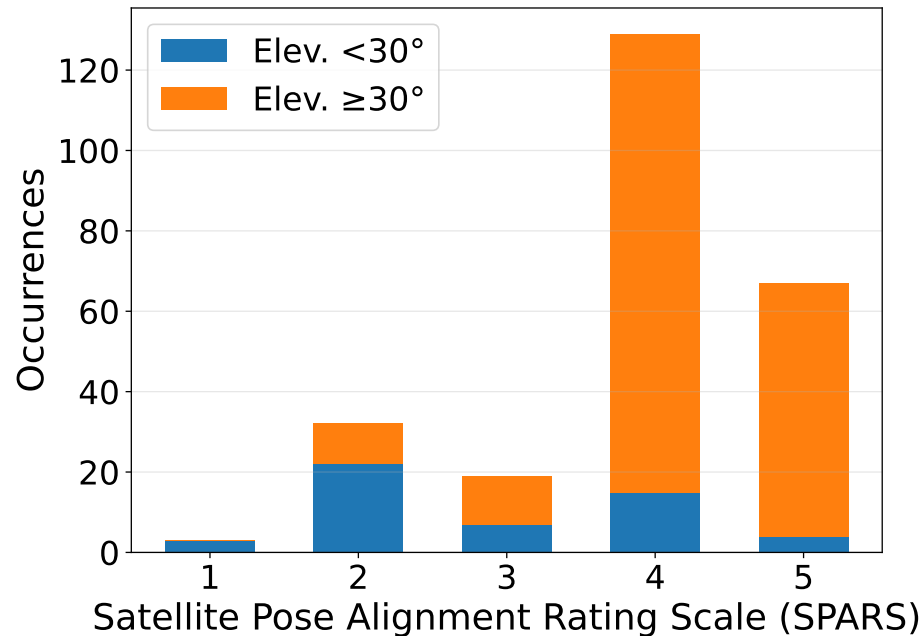


Figure 26. Histogram of SPARS evaluation categories for 250 real Seasat images scored by AFRL. Bars are stacked by target elevation $< 30^\circ$ vs. $\geq 30^\circ$ (model trained only for $\geq 30^\circ$).

Table 4. Qualitative evaluation results for 250 real Seasat test images using the SPARS metric, separated into elevation angles below and above the 30° training cutoff.

SPARS Level	All ($N = 250$)	Elev. $< 30^\circ$ ($N = 51$)	Elev. $\geq 30^\circ$ ($N = 199$)
SPARS-1 (Catastrophic Failure)	3 (1.2%)	3 (5.9%)	0 (0.0%)
SPARS-2 (Limited Success)	32 (12.8%)	22 (43.1%)	10 (5.0%)
SPARS-3 (Moderate Success)	19 (7.6%)	7 (13.7%)	12 (6.0%)
SPARS-4 (High-Confidence Match)	129 (51.6%)	15 (29.4%)	114 (57.3%)
SPARS-5 (Ground Truth Equiv.)	67 (26.8%)	4 (7.8%)	63 (31.7%)
Mean SPARS	3.90	2.90	4.16

3.1.3. Hubble Space Telescope (7DOF, Framewise Inference)

HST presents greater difficulty than Seasat due to strong symmetry and a solar array rotation. Although HST has up to 13DOF including solar arrays, aperture door, and high-gain antennae, it is simplified to 7DOF here based on image fidelity (antennae are unresolved) and the assumptions that the aperture door remains open and the solar arrays articulate in unison. A single published frame [49] yielded 2.6° rotation error and 11 cm translation error (160 nrad at the cited range); the symmetry-adjusted solar array angle error was 8° (Figure 27). On an unreleased 249-frame pass processed independently (no filtering), the mean body rotation error was 27.1° (median 6.4°), while mean image-plane translation error was 54 cm (median 41 cm). Restricting to elevation $>30^\circ$ did not meaningfully affect these statistics. Complete metrics are reported in Table 5. Given an estimated ground-truth label uncertainty of 1° – 5° and the small test set, minor performance variations should be interpreted conservatively.

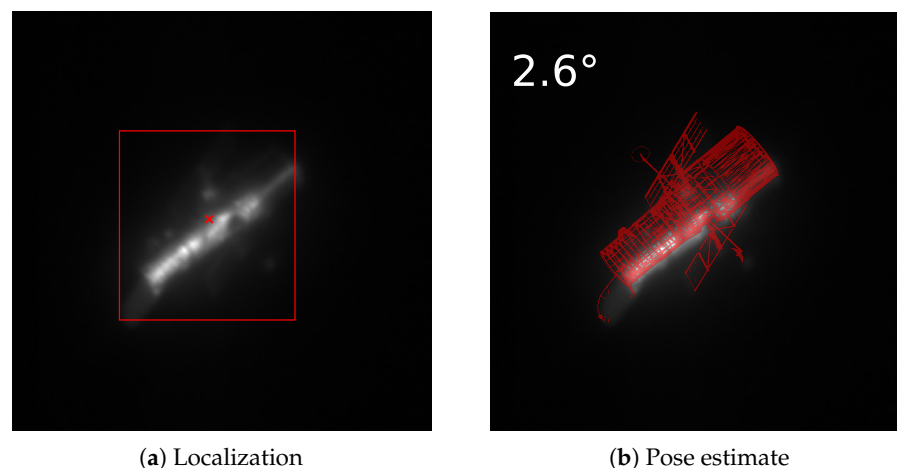


Figure 27. HST pose overlay for the published frame in [49]; predicted 7DOF pose (including solar array rotation) shown as a red wireframe overlay; body rotation error in degrees at upper left. Panel (a) shows the bounding box and object center output from the Coarse Localizer.

Table 5. Rotation, translation, and solar array angle error statistics for real HST test imagery. The Bennett et al. [49] image is a single published frame. The 249-frame video dataset was processed independently (no temporal context), with and without Kalman filtering. Solar array angle errors are adjusted for 180° symmetry where noted.

Dataset	Rotation Error (°)				Translation Error (cm)				Solar Array Error (°)	
	Mean	Median	Min	Max	Mean	Median	Min	Max	Mean	Max
Bennett et al.	2.6	–	–	–	11	–	–	–	8 ^a	–
249 Frame Video (No KF)	27.1	6.4	0.6	179.8	54	41	1	245	19 ^a	88 ^a
249 Frame Video (KF Applied)	5.5	4.1	0.2	23.2	48	40	0	236	8 ^a	23 ^a
Standard Deviation	3.9 ^b	–	–	–	38 ^b	–	–	–	7 ^b	–

^a Solar array angle errors adjusted for 180° symmetry. ^b Standard deviation for Kalman-filtered 249-frame video.

3.1.4. Hubble Space Telescope (Kalman Filtered)

Applying real-time Kalman filtering to the same 249-frame pass reduced mean body-rotation error from 27.1° to 5.5° (median 6.4°→4.1°). Mean image-plane translation error decreased from 54 cm to 51 cm (median 41 cm→40 cm). Mean solar array rotation error (symmetry-adjusted) improved from 19° to 8° (median 13°→7°). Complete metrics are reported in Table 5. Error traces are shown in Figure 28 (body) and Figure 29 (array).

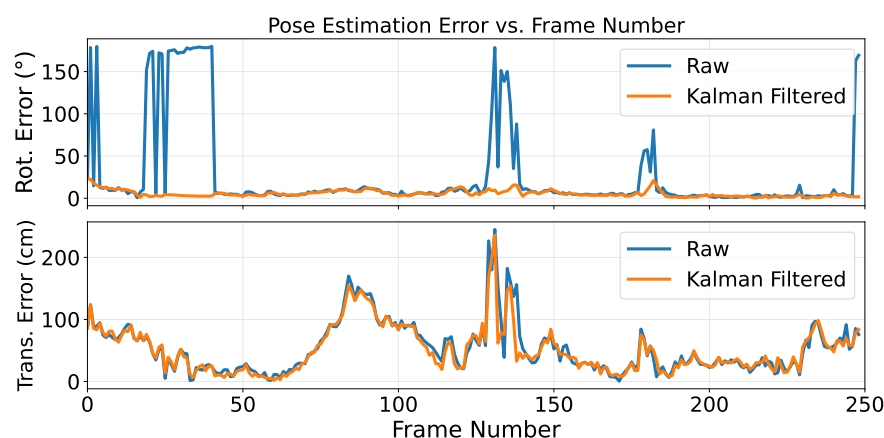


Figure 28. Per-frame pose errors across 249 real HST frames. **(Top)** body rotation error (deg). **(Bottom)** image-plane translation error (cm). Raw (blue) vs. Kalman-filtered (orange).

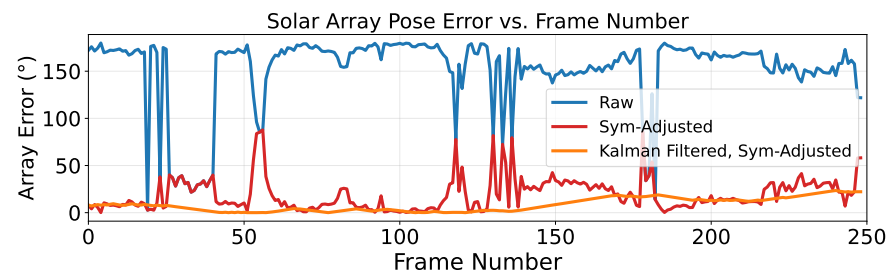


Figure 29. Per-frame solar array rotation error for the 249-frame HST pass. Raw (blue) exhibits 180° ambiguity; symmetry-adjusted (red) and Kalman-filtered (orange) curves show substantial reduction.

3.2. HFWO Synthetic Results and Image-Quality Dependence (DIRSIG + HCIPy)

This subsection consolidates pose model HFWO accuracy and its dependence on atmosphere/image quality. All numbers below use independent-frame inference unless otherwise stated. Summary tables for Seasat, ARGOS, and HST appear in Tables 6–8. Trends with AO-IQ are visualized in Figure 30. AO-IQ is calculated with the General Image-Quality Equation (GIQE) v3.0 [18,50,51].

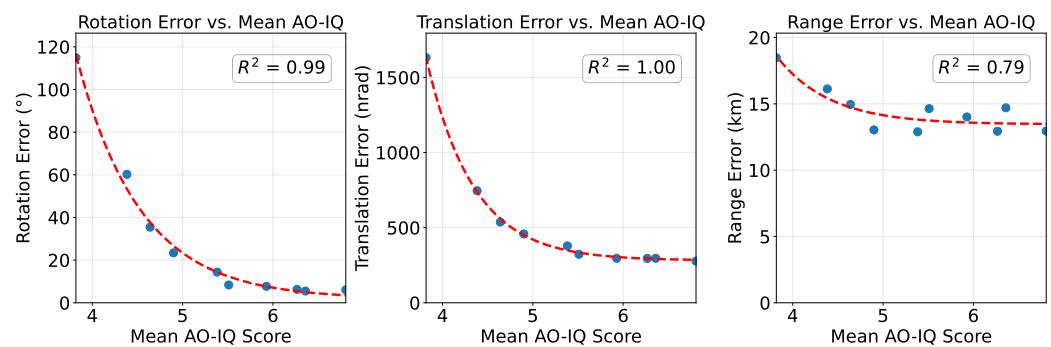


Figure 30. Seasat errors were averaged over each atmosphere and plotted as a function of mean adaptive optics image quality (AO-IQ) score for the test images resulting from that atmosphere (displayed as blue scatter plots). Exponential decay functions were fit to the data (displayed as red dashed curves), demonstrating high coefficients of determination.

Table 6. Image quality and pose estimation error metrics for Seasat as a function of r_0 . Higher r_0 values correspond to reduced atmospheric turbulence and improved image quality, as reflected by increasing AO-IQ and Signal-to-Noise Ratio (SNR). Pose estimation accuracy improves with image quality, showing reduced rotation, translation, and range errors. All values are means computed over 200 High Fidelity Wave Optics (HFWO) synthetic test images.

r_0 (cm)	AO-IQ	SNR	Rot. Err. (°)	Trans. Err. (nrad)	Trans. Err. (cm)	Range Err. (km)	Range Err. (%)
2.5	3.82	7.08	115	1630	170	18.5	1.79
3.5	4.38	9.23	60.2	746	77.0	16.1	1.56
4.0	4.64	10.8	35.5	537	57.2	15.0	1.45
4.5	4.90	12.2	23.4	458	47.8	13.0	1.26
5.0	5.38	13.5	14.4	379	39.6	12.9	1.25
6.0	5.51	15.8	8.35	323	34.0	14.6	1.42
7.0	5.93	17.8	7.69	296	30.3	14.0	1.36
7.5	6.27	18.8	6.28	295	31.0	12.9	1.25
8.5	6.36	20.5	5.51	296	31.1	14.7	1.43
10.0	6.81	22.7	6.09	277	28.6	12.9	1.26

Table 7. Image quality and pose estimation error metrics as a function of r_0 for the satellite ARGOS. Pose estimation accuracy improves with image quality, showing reduced rotation, translation, and range errors. All values are means computed over 200 HFWO test images. Rotation errors are presented both in raw and symmetry-adjusted form.

r_0 (cm)	AO-IQ	SNR	Rot. (°)	Rot. Adj. (°)	Trans. (nrad)	Trans. (cm)	Range (km)	Range (%)
5	5.42	12.2	77.5	33.0	472	53.7	17.3	1.55
6	5.99	14.7	66.6	20.1	350	40.3	15.9	1.42
7	5.98	17.2	61.3	16.6	331	37.7	15.6	1.40
10	7.04	22.6	56.8	11.9	302	34.5	15.8	1.42

Table 8. Image quality and pose estimation error metrics as a function of r_0 for the HST with 7DOF total (including solar array rotation). Means over 200 test images; rotation errors also shown in symmetry-adjusted form.

r_0 (cm)	AO-IQ	SNR	Rot. (°)	Rot. Adj. (°)	Trans. (nrad)	Trans. (cm)	Range (km)	Range (%)	Array (°)	Array Adj. (°)
5	5.85	15.68	37.3	20.3	1262	92.2	10.9	1.48	103.3	21.4
6	6.22	18.33	26.5	18.0	1186	86.6	11.2	1.52	102.1	14.7
7	6.64	20.31	24.8	13.9	1170	85.4	10.4	1.41	105.1	15.6
10	7.21	25.10	20.1	11.1	1100	80.1	10.9	1.49	107.6	12.1

A comment is warranted on the apparent discrepancy between qualitative results showing excellent silhouette alignment (Figure 16) and quantitative errors (Figure 31b). The large rotational errors are caused by 180° ambiguities inherent to geometrically symmetric satellites like ARGOS and HST, a weakness intentionally probed by the Sim2Sim domain gap (Section 2.4). This challenge also manifests for real imagery (Figures 28 and 29). While higher-resolution imagery or use of context clues (such as assuming a nominal sun-tracking orientation for the solar arrays) can mitigate this, time-series analysis provides another solution. For instance, Kalman filtering resolved the main body ambiguity for HST, though it was insufficient for the more difficult solar array symmetry. These ambiguities are challenging for human analysts as well and are a key direction for future research, with a full discussion available in [18].

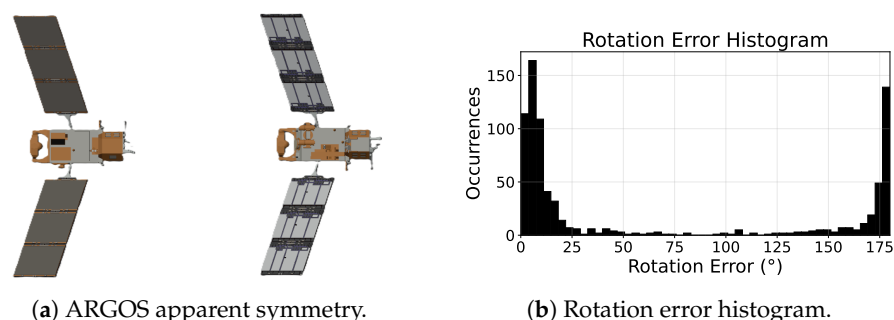


Figure 31. (a) Two views of opposing sides of the ARGOS satellite CAD model. With the exception of small components not clearly visible under reduced contrast at high spatial frequencies, the silhouettes are nearly identical. (b) Histogram of rotation errors (in degrees) for ARGOS, with a significant cluster near 180° indicating symmetry as a primary source of error.

3.2.1. Seasat

For $r_0 = 6$ cm the mean rotation error was 8.4° and the mean two-axis image-plane translation error was 323 nrad (34 cm) (Table 6), in line with the model performance on

real data. Accuracy improves nearly monotonically with AO-IQ (Figure 30 and Table 6). For equivalent atmospheres the Seasat model provides improved accuracy relative to the ARGOS (smaller, highly symmetric) and HST (7DOF, symmetric) models, illustrating that Seasat's large size and strong geometric asymmetry make it an ideal target for pose estimation.

3.2.2. ARGOS

ARGOS presents strong apparent symmetry (Figure 31a), making it a much more challenging test target than Seasat (and perhaps more representative of modern LEO satellite geometries). For $r_0 = 6$ cm the raw mean rotation error was 66.6° and the mean image-plane translation error was 350 nrad (40 cm).

To account for the target's geometric symmetry, rotation errors are reported in both raw form (e_q) and as a 3D symmetry-adjusted metric (e_{sym}). Unlike the single-angle scalar adjustment used for solar arrays in Section 2.5.6, adjusting a full 3D orientation requires evaluating the error relative to the satellite's physical axis of symmetry. A near- 180° orientation flip produces a visually equivalent silhouette only when the rotation occurs about an actual symmetry axis of the target. For ARGOS, 180° rotational symmetry exists about the primary longitudinal axis of the bus (Figure 31a). The 3D symmetry-adjusted rotation error e_{sym} is calculated using the quaternion error metric e_q defined in Equation (5):

$$e_{\text{sym}} = \min(e_q(\hat{q}, q), e_q(\hat{q}, q \otimes q_{x,180^\circ})), \quad (6)$$

where $\hat{q}$ is the predicted quaternion, q is the ground-truth quaternion, $q_{x,180^\circ}$ represents a 180° rotation about the CAD X-axis, and $\otimes$ denotes quaternion composition under Shuster's convention.

Symmetry-adjustment reduced the mean rotation error for $r_0 = 6$ cm from 66.6° to 20.1° (Table 7). Qualitative overlays show excellent silhouette alignment despite large raw rotation errors (Figure 16); the error histogram concentrates near $0^\circ/180^\circ$ (Figure 31b). The Sim2Sim domain gap presented a significant challenge: replacing DIRSIG object planes with Blender object planes (same HCIPy PSFs, $r_0 = 6$ cm) reduced raw rotation errors from 66.6° to 17.2° and translation error from 350 nrad to 276 nrad.

3.2.3. Hubble Space Telescope (7DOF)

At $r_0 = 6$ cm the HST achieved 26.5° mean body-rotation error, 18.0° symmetry-adjusted body-rotation error (calculated with Equation (6)), 87 cm mean image-plane translation error, and 14.7° mean symmetry-adjusted solar array angle error (Table 8). Overlays again show good silhouette alignment with symmetry-driven ambiguity (Figures 17 and 32a); the rotation-error histogram exhibits peaks near $0^\circ/180^\circ$ (Figure 32b). Errors decrease with r_0 and AO-IQ.

3.2.4. Dependence on Atmosphere and Image Quality

Across satellites, pose accuracy increases with r_0 and AO-IQ. After averaging errors over pose and illumination for Seasat across ten atmospheric bins ($r_0 \in [2.5, 10.0]$ cm), exponential decay functions fit to mean error vs. mean AO-IQ achieve $R^2 \approx 0.99$ (rotation) and $R^2 \approx 1.00$ (translation) (Figure 30). However, these fits represent an exploratory empirical correlation evaluated on the same atmosphere-averaged synthetic data points used for fitting. Because the relationship was not validated on a held-out test set or real imagery, AO-IQ should be interpreted as an initial exploratory correlate rather than a universally validated predictor. Fit parameters carry inherent uncertainty, and the exact functional form is expected to vary across different satellite geometries, AO systems, and vision models.

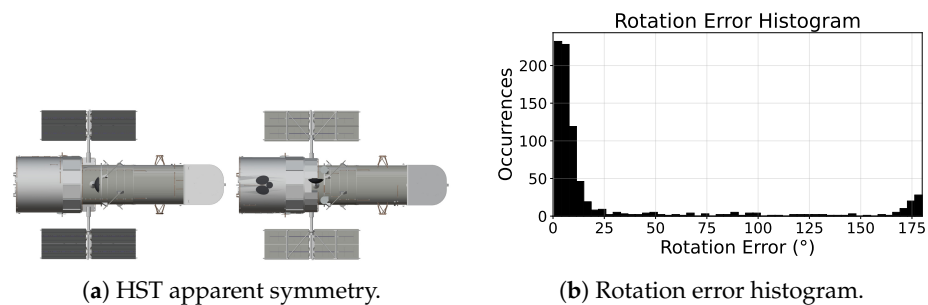


Figure 32. (a) Two views of opposing sides of the HST satellite CAD model. As with ARGOS, the silhouettes are highly similar. (b) Histogram of HST rotation errors shows clustering near 180° , indicating apparent symmetry as a key error source, albeit less dominant than for ARGOS (see Figure 31b).

3.3. Pose Accuracy vs. Object Pose

To isolate orientation effects, a single AO PSF ($r_0 = 5$ cm) and realistic noise were applied (via HCIPy) to all 11,000 pristine Seasat LFID test renders, holding image quality fixed. To determine if the model struggles with specific viewing angles, we grouped the images into distinct orientation categories. Using ground-truth rotations, unsupervised k -medoids clustering (quaternion distance) partitioned the 3DOF orientation space into 110 pose classes (mean intraclass quaternion distance 34.3° , interclass 127.4° ; ~ 100 images/class).

Analyzing these classes revealed that estimation accuracy depends heavily on the target's pose. Class-mean rotation errors spanned 3.1° – 26.9° (global mean 8.3°), with an interclass standard deviation of 4.6° (coefficient of variation (CV) = 0.55), indicating substantial pose-dependent difficulty, though intraclass variability remained high (mean intraclass std dev 23.6°). Translation estimation showed a more moderate dependence on pose (class means 21–72 cm; global mean 33 cm; CV = 0.27; mean intraclass std dev 28 cm), while range estimation exhibited the weakest dependence (class means 10.2–21.5 km; global mean 13.1 km; CV = 0.16; mean intraclass std dev 9.4 km).

Physically, accuracy was poorest when Seasat was oriented with the solar arrays closer to the sensor, obscuring portions of the satellite body behind them. In contrast, accuracy was highest with Seasat oriented with the synthetic-aperture radar (SAR) panels towards the top of the image, the solar arrays towards the bottom of the image, and the boom antennas projecting towards the right of the image, as shown in Figure 18.

3.4. Pose Accuracy vs. Illumination Direction

Using the same 11,000-image set, illumination orientations (in image coordinates: +X right, +Y up, +Z out of the image) were clustered into 110 classes. Rotation error exhibited a strong dependence on lighting: global mean 8.42° with class means from 2.89° to 38.3° (std dev of class means 7.21° , CV = 0.86; mean intraclass std dev 23.1°). Translation showed a modest dependence (means 23–57 cm; global mean 33 cm; CV = 0.20), and range was weakly affected (means 9.70–16.7 km; global mean 13.1 km; CV = 0.12). A significant correlation was found between rotation error and the illumination Z-component (Pearson $r = +0.215$, $p < 10^{-100}$): back-illumination ($Z \rightarrow +1$) increased error, while front-illumination ($Z \rightarrow -1$) reduced error, as shown in Figure 19. Further analyses are provided in [18].

3.5. Pose Accuracy vs. CAD Fidelity

Four independent Seasat training sets captured geometry/material realism at increasing fidelity: (i) 100-triangle, single material; (ii) 729-triangle, 15 materials (decimated); (iii) full-fidelity 34k-triangle, 17 materials; (iv) full-fidelity with randomized material augmentation (Figure 33). Each Coarse Localizer/PoseNet was trained identically and evalu-

ated on 2000 HFWO test images. Mean rotation/translation errors improved monotonically with fidelity: $77.8^\circ/689$ nrad (100 tri) $\rightarrow$ $54.9^\circ/617$ nrad (729 tri) $\rightarrow$ $46.5^\circ/569$ nrad (34k-tri) $\rightarrow$ $42.4^\circ/537$ nrad (34k+mat-aug). The model degrades gracefully under reduced CAD realism. Additional detail is available in [18].

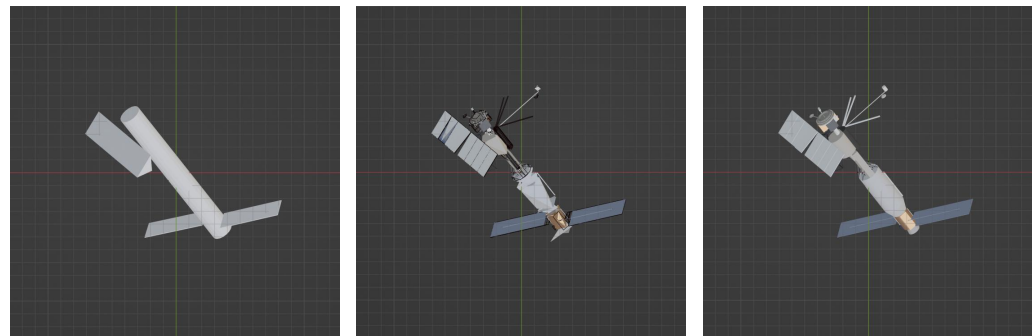


Figure 33. The single-material 100-triangle Seasat CAD model, composed of a cylinder, triangular prism, and two rectangular prisms is at (left). The 15-material 729-triangle Seasat CAD model created by applying Blender’s decimate modifier to the original model is at (center). The 17-material 34k-triangle full-fidelity Seasat CAD model is at (right).

3.6. Pose Accuracy vs. Training Set Size

The full 264k-sample training set (88k unique poses, $3\times$ randomized augmentations) was subsampled to 16k, 32k, 64k, 88k, and 176k; the pose model was retrained per subset while keeping the localizer fixed (trained on 264k). All models were evaluated on the same 2000-image HFWO test set used in Section 3.5. Mean errors (averaged across 10 atmospheres) improved with dataset size: rotation decreased from 73.4° (16k) to 42.4° (264k), translation from 867 nrad to 537 nrad, and range from 16.2 km to 14.5 km. An exponential decay fit to rotation error vs. dataset size achieved $R^2 = 0.96$; extrapolation to 1.264 M samples projects $\sim 30^\circ$ mean rotation error, indicating continual (but diminishing) improvement. Additional details are provided in [18].

3.7. Human Analyst Comparison

A preliminary benchmark was conducted on 200 Seasat HFWO test images at $r_0 = 5$ cm (shorter simulated integration time, $\sim 50\%$ lower SNR than Section 3.2.1). Frames were processed independently without temporal context by a single expert human annotator (the first author, using a custom graphical user interface (GUI) detailed in [18]) and the pose model to provide a limited evaluation of manual labeling consistency and speed as a single point of comparison. Cumulative error distributions are shown in Figure 34.

Rotation: The mean human rotation error was 66.6° , whereas the model achieved 34.4° , outperforming the analyst on 141/200 images. Rotation errors ranged from 0.6° to 180.0° for the human and from 0.3° to 177.0° for the model.

Translation: The mean absolute two-axis image plane angular translation error was 572 nrad (59.3 cm) for manual labeling and 473 nrad (48.3 cm) for the model, with the model outperforming the analyst on 116/200 images. Translation errors ranged from 54 to 3812 nrad for the human and from 27 to 2003 nrad for the model. Range was not manually labeled, and the same range estimate provided to the model for inference was used to render the labeling GUI.

Throughput: Manual labeling required 6.27 h in total (mean 1 min 43 s per image; range 23 s to 8 min 37 s). Model inference at 7.1 Hz on a RTX 3060 GPU processed the same set in 28 s ($\sim 800\times$ faster).

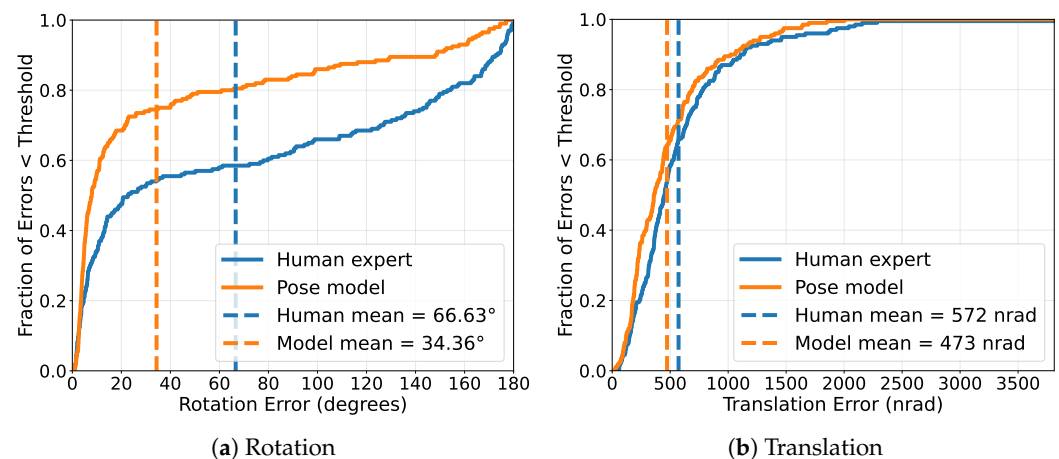


Figure 34. Cumulative fractional distributions of rotation (a) and two-axis apparent angular translation (b) errors comparing performance of a single human labeler to the computer vision pose model. In this highly limited and preliminary experiment, the vision model demonstrated lower error.

3.8. Approach Practicality, Runtime, and Hardware Requirements

A key aim of this research effort was operational practicality. Training data were synthesized with Blender (v3.6.0) and Python (v3.9.16) tooling (RTX 3060 GPU, 6 GB VRAM; 32 GB RAM). Model training and augmentation ran on a single NVIDIA A100 (40 GB VRAM) [44]. Rendering 110k 256×256 grayscale images plus paired front-lit views required ~ 10.5 h; rendering $286,000 \times 64 \times 64 \times 3 M_{XYZ}$ added ~ 2.5 h. Data loading/preprocessing and augmentations contributed ~ 3.3 h total (CPU). A100 training required ~ 8.5 h for the Coarse Localizer, ~ 11 h for the complete PoseNet backbone, and ~ 3.5 h for the pose head. With early stopping (50-epoch caps; patience 10), the full process (including data generation) required < 40 h. Many stages (rendering, augmentation, RoI crops) are trivially parallelizable and could be distributed to reduce wall time further. Pretrained backbones or transfer learning were not used but could likely shorten training.

The storage footprint per satellite was ~ 25 GB (lossless compression), dominated by ~ 18 GB for 286k augmented 256×256 16-bit images; weights were ~ 100 MB (Coarse Localizer) and ~ 350 MB (pose network). At inference, the end-to-end pipeline (localizer $\rightarrow$ RoI crop $\rightarrow$ PoseNet $\rightarrow$ optional AKF) sustained ~ 7.1 Hz in real-time streaming mode on the RTX 3060, ingesting frames sequentially with batch size = 1.

4. Discussion

This work delivers and validates the first practical end-to-end pipeline for real-time, ≥ 6 DOF satellite pose estimation from ground-based AO imagery, trained exclusively on synthetic data. The system generalizes well to real imagery: Seasat achieves $\sim 5^\circ$ mean rotation error and 21 cm mean image-plane translation error without temporal filtering, and a full HST pass improves from 27.1° framewise error to 5.5° with Kalman filtering. A qualitative evaluation of 199 additional Seasat frames at elevation $\geq 30^\circ$ rated 88.9% of predictions as “ground-truth equivalent” or “high confidence” (78.4% across all 250 frames; Section 3.1.2). On a high-fidelity synthetic test set (HFWO, $r_0 = 6$ cm), the model attains 8.4° mean rotation and 34 cm mean image-plane translation error, consistent with real-image results. Operational viability is confirmed: end-to-end inference runs at 7.1 Hz on consumer hardware and achieved 48.4% lower mean rotation error than a single human labeler on a limited 200-image benchmark set while operating $\sim 800\times$ faster (Figure 34). Finally, we propose an exploratory empirical correlate linking pose estimate accuracy to image quality: under controlled synthetic conditions ($N = 10 r_0$ bins) and without a held-out test set, the scalar image-quality proxy AO-IQ appears to account for most of the

variance in rotation and translation error for Seasat after averaging over object pose and illumination direction (Figure 30).

4.1. Interpretation of Results in Context

This work advances beyond prior art ([3,4]) by enabling full ≥ 6 DOF pose inference under realistic imaging conditions, validated on real imagery. Key, application-specific architectural adaptations from GDR-Net include:

- Serial backbone: four smaller U-Nets (reconstruction $\rightarrow$ relighting $\rightarrow$ segmentation $\rightarrow$ pose) stabilize training and improve robustness to degraded inputs, while remaining single-GPU trainable (Section 3.8, Figure 5).
- Physical priors: injecting illumination direction and TLE-derived range estimates improves accuracy.
- Coupled training: training the PoseNet on the Coarse Localizer outputs reduces tail errors from poor image quality (29% improvement in 90th-percentile error, Figure 23).

Additional adaptations include target range reparameterization (suitable for ranges of ~ 1000 km), auxiliary localization inputs at the feature level, main body pose-aware branched regression heads for articulating components, and a tailored data augmentation pipeline randomized over blur, noise, and other SDA AO-specific degradations ([19], Section 2.3.4). These design choices enable performant generalization to unseen AO data and extension to 7DOF pose estimation (Figure 22).

Direct horizontal benchmarking against established space-based RPO pose estimation algorithms (e.g., SPEED+ [13]) is limited by fundamental domain gaps. The space-to-space RPO models operate on high-resolution (1920×1200) near diffraction-limited imagery at close ranges (2–10 m) with significant high spatial frequency structure, whereas ground-based AO pose estimation processes RoI inputs (64×64 pixels) at LEO slant ranges (~ 1000 km) subject to residual AO blur and low SNR. Because no prior baseline exists for 6DOF pose estimation from ground-based AO SDA imagery, this work establishes the initial performance baseline for this regime, evaluated against a human labeler (Figure 34) and generalist vision models (Appendix A). Direct horizontal comparison across approaches will require the creation of standardized ground-based AO benchmark datasets, which is a future work priority.

4.2. Key Challenges and Limitations

Remaining limitations include both algorithmic and practical factors:

- Symmetry ambiguities: Symmetric geometries (ARGOS, HST) drive residual error. On ARGOS at $r_0 = 6$ cm for HFWO testing, raw mean rotation error is 66.6° , dropping to 20.1° after symmetry adjustment (Table 7, Figure 16). Kalman filtering has the ability to substantially mitigate geometric ambiguity (Figure 28).
- Dependence on CAD fidelity: Pose accuracy degrades gracefully with decreased CAD fidelity (Section 3.5). Accurate CAD models are still required at train time. Ongoing work in automated 3D mesh generation and model-free pose estimation ([52,53]) is relevant.
- 7DOF limits: For HST, farthest point sampling semantic segmentation U-Net output labels were insufficient to guide learning for flat and highly symmetric components (solar array front/back). Explicit front/back surface symmetry-breaking supervision is a recommended future direction (Figure 29). Alternatively, known illumination direction can be used to break symmetry for solar arrays.
- Slant range validation for real data: Predicted slant range (Z) accuracy was quantified for HFWO synthetic datasets but could not be quantitatively evaluated for real imagery due to the absence of ground-truth slant range and sensor IFOV.

- *n*-DOF scaling and accuracy degradation: While the PoseNet architecture is structurally extensible to *n*-DOF (Section 2.5.6), accuracy is expected to decrease with increasing *n*. As *n* increases, M_{SRA} must allocate fewer pixels and classes to each individual articulating component, reducing the ability to guide regression. Furthermore, branching multiple regression heads off a single shared backbone for independent articulations may introduce additional gradient interference during training. Consequently, while the architecture structurally supports $n > 7$, achieving useful accuracy for high *n* targets will likely require new advancements.
- Target selection and geometric diversity: The evaluation targets (Seasat, HST, and ARGOS) were deliberately selected to span three general satellite structural classes: large asymmetric (Seasat), large symmetric with articulation (HST), and small symmetric (ARGOS). This expands upon single-target evaluations in prior satellite pose estimation literature ([4,14]). However, validating the pipeline across a broad diversity of satellite geometries, sizes, and orbital regimes remains necessary to fully characterize performance.
- Scope of architectural ablations: System design choices were ablated for key targeted mechanisms. Specifically, coupled training for Coarse Localizer and PoseNet (Figure 23), CAD model fidelity (Section 3.5), and training set size (Section 3.6). Additional ablations on 3DOF pose representation and output activation function and neural network parameter count are reported in [18] Chapter 3. However, a systematic combinatorial ablation for each U-Net remains to be conducted given limited resource availability and prioritization and represents a limitation of the current work.

While the pipeline demonstrates robust performance under nominal conditions, compiling the extremes from our evaluations quantitatively defines the approach's operational boundaries. System performance appears to degrade gracefully in most cases:

- Atmospheric and image quality limits: As established in Table 6, operation at $r_0 \leq 4.0$ cm ($AO-IQ \leq 4.64$) marks a boundary for reliable pose recovery, with mean rotation errors exceeding 35° on the HFWO Seasat dataset as high spatial frequency structural details are lost to residual AO blur. As noted, Seasat's large size and asymmetric geometry present a best-case bound for performance. Target elevation angle was constrained to $\geq 30^\circ$ in training and testing because increased atmospheric path length for low elevations degrades typical image quality to below $AO-IQ \sim 5$.
- Illumination bounds: Performance is strongly correlated with the illumination vector's Z-component (Section 3.4). Optimal front-illumination yields rotation errors as low as 2.89° , whereas deep back-illumination establishes the upper operational error boundary at 38.3° due to the loss of visible features in deep shadows. Again, Seasat presents a best-case performance bound.
- Attitude and obscuration limits: As shown in Section 3.3, target self-obscuration correlated with pose accuracy. Poses where primary structural features (e.g., the satellite bus) are heavily self-occluded along the line of sight degrade mean rotation accuracy to 26.9° for Seasat, compared to 3.1° under optimal viewing geometry. Target image-plane translation was constrained such that the full extent of the target was maintained within the FOV. Pre-existing operational AO tracking systems easily maintain the target within the FOV, and cropping the target is expected to significantly reduce accuracy.
- Geometric fidelity: The system requires synthetic training data containing geometry that is closely representative of the true target geometry to function. Reducing CAD fidelity to a minimally recognizable version of the object increased Seasat's rotation error to 77.8° from 42.4° for the highest fidelity CAD representation on the same HFWO Seasat test set (Section 3.5).

- Slant range and spatial resolution limits: Target slant range dictates image-space resolution for a given sensor IFOV. Evaluation across slant ranges from 735 km to 1300 km for Seasat found that rotation error did not correlate with space sample distance (SSD) ($R^2 \approx 0.01$), a direct proxy for slant range. Further study is required to establish a SSD or slant range threshold for which target extent becomes insufficient to resolve key structural features for pose regression.

4.3. Implications for SDA

The pipeline provides a validated approach for real-time monitoring and enhanced operator situational awareness as well as for post-processing imagery archives into pose histories. Intermediate model outputs (M_{SRA} , M_{XYZ} , etc.) require negligible overhead and support additional analysis. Despite advances in generalist VLMs and zero-shot depth estimators, preliminary testing (Appendix A) indicates these models fail on unconstrained pose and depth estimation tasks on SDA imagery unless aided by internal tools (e.g., image centroiding using Python during chain-of-thought) and restricted to highly simplified settings. Specialized, domain-adapted models remain necessary for practical SDA utility at this time.

4.4. Future Research Directions

Priority areas for further research include:

- Sim2Real and domain adaptation: Explore style/domain transfer (NST, StyleGAN-based randomization), adversarial adaptation (DANN), and self-training on unlabeled AO sequences; evaluate domain generalization frameworks that perform well on SPEC2021 [13,54].
- Improved symmetry handling: Evaluate ARGOS pose model on real imagery and devise architectures and approaches more robust against strong geometric symmetry.
- Temporal modeling: Replace test-time filtering with Recurrent Neural Networks (RNNs) or Transformers trained on AO video to learn dynamics; investigate physics-informed priors [55–57].
- Test-time refinement: A novel render-and-compare procedure [18] yields modest gains on high-quality frames ($4.5^\circ \rightarrow 3.4^\circ$ mean rotation error, 24 cm $\rightarrow$ 19 cm mean translation error at $r_0 = 8.5$ cm) but is ineffective on marginal imagery (where it would add the most value) and computationally impractical (5–30 s/frame). Improved test-time refinement methods could prove useful.
- Hybrid geometric learning: Fuse learned dense correspondence regression with PnP, following top SPEC2021 pipelines [58,59]; revisit off-the-shelf localizers (e.g., Faster R-CNN [37]).
- Efficiency: Enable multi-target model training via satellite ID injection, implement parallel data generation, and leverage transfer learning across satellites. Improve range handling by treating TLE-range and sensor IFOV as ground truth and resizing CAD geometry accordingly. Explore more advanced pretrained backbone architectures.
- SPARS qualitative scale validation: Validate the SPARS framework across multiple independent analysts to establish intra-rater repeatability and inter-rater agreement metrics, addressing the single-rater limitation of the current evaluation and confirming its operational utility.
- AO-IQ validation: Perform formal held-out validation and uncertainty quantification for the AO-IQ metric across diverse test sets.
- Quantitative real-data range validation: Evaluate the pipeline across thousands of real image frames with known sensor IFOV and accurate TLE-calculated target ranges to quantitatively validate slant range error.

- High-DOF architectures: Develop and validate architectures and training methods to mitigate the expected accuracy degradation when scaling pose estimation to $n > 7$ degrees of freedom.
- Generalized model benchmarking: Propose a public SDA or pose estimation benchmark to track VLM vs. specialized approach performance.
- Ground-based AO pose benchmarking: Establish public ground-based AO dataset benchmarks (analogous to SPEED+ for space-based RPO) to enable direct cross-algorithm comparisons across emerging ground-based pose models.
- Fleet and orbital regime scaling: Validate performance across a broader diversity of satellite geometries, sizes, and orbits beyond the three classes evaluated here.

5. Conclusions

We present the first practical, real-time (7.1 Hz) system for estimating ≥ 6 DOF satellite pose from resolved ground-based AO imagery, trained purely on synthetic data. The model generalizes to real targets: Seasat achieves $\sim 5^\circ$ mean rotation and 21 cm mean image-plane translation error without temporal filtering; HST achieves 27.1° framewise error that falls to 5.5° with Kalman filtering. Compared to a single human labeler in a preliminary and limited experiment, the system is both faster ($\sim 800\times$) and demonstrates lower mean rotation error on the evaluated subset. We quantify accuracy as a function of image quality, pose, and illumination; AO-IQ explains $\sim 99\%$ of atmosphere-driven variance for Seasat, enabling actionable performance forecasting. We extend the framework to 7DOF and demonstrate it for HST, and document limitations due to satellite geometric symmetry and CAD fidelity. Finally, we suggest with a preliminary experiment in Appendix A that contemporary generalized VLMs are not yet viable substitutes for domain-specific SDA vision models.

This work establishes an operationally relevant baseline for automated satellite characterization from resolved imagery. By moving beyond theoretical exploration to a validated, high-performance system, it provides a scalable methodology for exploiting SDA sensor data at speed and at scale. As LEO becomes increasingly congested, such capabilities will be central to sustained situational awareness and space traffic management.

Author Contributions: Conceptualization, T.D., M.G., D.W., J.F. and D.M.; methodology, T.D., M.G. and D.W.; software, T.D. and D.F.; validation, T.D., M.G. and D.F.; formal analysis, T.D. and M.G.; investigation, T.D., D.F. and M.G.; writing—original draft preparation, T.D.; writing—review and editing, T.D., D.F., J.F., D.W., D.M. and M.G.; supervision, J.F., D.W., D.M. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2026-2757. This research was funded by the Department of Defense and Air Force Research Laboratory. The views expressed are those of the author and do not reflect the official policy or position of the US Air Force, US Space Force, Department of Defense, or the US Government.

Institutional Review Board Statement: Ethical review and approval were waived for this study. The work involved authors performing performance benchmarks and qualitative scoring of image data, a common practice in the field. The tasks posed no participant risk beyond that of normal computer use.

Informed Consent Statement: Not applicable. The only human subjects involved were the authors.

Data Availability Statement: The datasets presented in this article are not readily available because access is restricted due to Department of Defense and Air Force Research Laboratory policies. Requests to access the datasets should be directed to the corresponding author.

Acknowledgments: The authors would like to thank Dimah Dera, James Albano, and Linwei Wang of the Rochester Institute of Technology for their valuable guidance and expertise. We also thank our colleagues at the Air Force Research Laboratory (AFRL) for their technical expertise and many helpful discussions. The authors acknowledge Research Computing at the Rochester Institute of Technology for providing computational resources and support that have contributed to the research results reported in this publication [44]. Approved for public release; distribution is unlimited. Public Affairs release approval #AFRL-2026-2757.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Evaluation of Generalized Models for SDA

We assessed several large, generalized models on SDA tasks: OpenAI and Google Gemini VLMs for monocular pose and Depth Anything V2 for monocular depth. Tests used pristine orthographic renders (512×512 px) and required pose estimates relative to a reference view.

Appendix A.1. Vision-Language Models

OpenAI's GPT-4o (gpt-4o-2024-11-20): performed poorly on rotation and translation: (i) relative 3DOF rotation from two views: 0/5 correct across two setups; (ii) 1DOF rotation classification with 10 classes: 3/30 correct ($\approx$ chance); (iii) 1DOF translations along X or Z: all incorrect. A heavily simplified 1DOF Z-axis rotation was answered correctly 2/3 times, with one success occurring only after corrective feedback. A March 2025 GPT-4o update [60] improved a single simplified case: with baseline/target embedded as red/blue channels, the model identified the axis and direction of 1DOF X translation in 5/5 trials, though magnitudes varied without a scale reference. Prompts, input images, and complete experimental procedures are provided in [18].

OpenAI's o4-mini-high (April 2025): showed materially better but still limited capability [61,62]. With pixel scale provided, it recovered 1DOF X translations with correct axis/direction in all trials and near-metric accuracy (red/blue overlay: $|\Delta X| = 5.7\text{--}6.02$ m for a 5 m truth; separate images: 5.00 m in all three trials). For 1DOF Z translation (separate images), it correctly inferred a twofold distance change (3/3). On the 10-class 1DOF rotation benchmark it reached 13/30 correct (43.3%), with strong Z-axis (image-plane rotation) results (7/9) but weak X/Y (out-of-plane rotation) (2/9 and 1/9). A simple 6DOF task with multi-view context remained unsolved (three incorrect attempts).

Google's Gemini Pro 3.1 Preview (March 2026): Ref. [63] continued the trend of improved performance, achieving perfect performance on the 1DOF Z translation and 1DOF X translation (separate images) tasks. Notably, while o4-mini-high also performed perfectly on these tasks it did so with extensive use of Python for calculations (such as centroiding) within the chain-of-thought, despite being instructed not to in the prompt. All tool use was explicitly disabled for Gemini Pro 3.1 Preview, so performance was purely the result of the multimodal model ingesting and operating on tokenized images. Gemini Pro 3.1 Preview also achieved a perfect 30/30 score on the 10-class 1DOF rotation benchmark. The prompts were modified, appending "The first image is the baseline pose", as Gemini Pro 3.1 Preview does not access the filenames of uploaded images and it was important that the model correctly distinguish the baseline reference pose from the target pose so that the predicted transformations were not inverted. The model consistently identified transformations about the X, Y, and Z axes, suggesting a more robust internal representation of 3D geometry. Inference times (driven mainly by number of reasoning iterations) varied significantly with apparent task complexity, from under 10 s to over two minutes. The model's capabilities with degraded imagery or on more challenging, unconstrained 6DOF pose estimation tasks remain to be tested. In a highly simplified 6DOF

pose test Gemini Pro 3.1 Preview improved on o4-mini-high with 3/3 correct rotations and 0/3 correct translations (o4-mini-high had zero correct rotations or translations). See [18] Section 9.1.8, “6DOF Pose Estimation with Context” for complete details.

While VLMs are progressing rapidly, they still require tens to hundreds of seconds of reasoning and remain limited to pose problems that are narrowly constrained with pristine data. This investigation is not intended as a comprehensive benchmark, and many VLMs were not evaluated (including OpenAI’s latest model), as that is outside the scope of this work. Rather, these results highlight the clear trend of improvement (a trend also reflected in broader industry benchmarks evaluating general spatial and object understanding [64]) and motivate the need for a formal, standardized pose estimation benchmark for generalized models.

Appendix A.2. General-Purpose Monocular Depth

Depth Anything V2 [65], trained at web scale (including ImageNet-21k classes such as satellite/space_station [66]), produces credible depth on everyday scenes but failed to recover meaningful structure on simulated SDA imagery of Seasat. A domain-specific depth model (the PoseNet backbone) performed substantially better (Figure A1) [18]. Given that these test images were in-distribution for PoseNet, its superior performance was expected. However, the severity of Depth Anything V2’s failure was unexpected given its strong cross-domain generalization and a massive training set of 595k synthetic images and ~62 M pseudo-labeled real images.

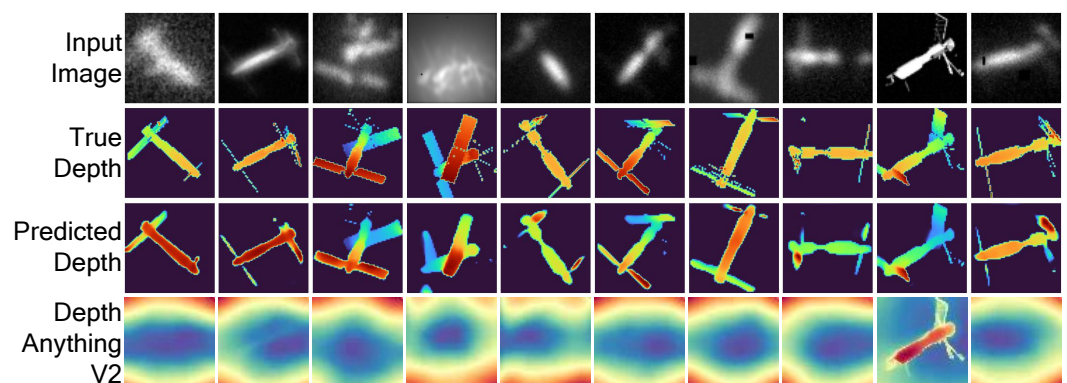


Figure A1. The top row shows ten simulated test images, the second row shows corresponding ground truth depth maps, the third row shows predicted depth outputs from a custom model trained on synthetic Space Domain Awareness (SDA) imagery, and the final row shows predicted depth outputs from Depth Anything V2 [65]. All depth maps use the turbo colormap (red = near, purple = far) with per-image min-max normalization (linear contrast stretching).

Appendix A.3. Takeaway

Generalized models did not provide reliable pose or depth estimation for SDA without task-specific conditioning. Limited successes occurred only in highly simplified settings with explicit pixel scale or strong priors. While recent industry evaluations demonstrate that frontier VLMs are achieving high accuracy in 2D visual understanding (such as object detection, identification, and counting [64]) metric 3D pose recovery on unconstrained SDA imagery remains an unaddressed challenge for generalist models. These findings highlight the distinction between 2D scene understanding and fine 3D estimation, reinforcing the need for domain adaptation and standardized pose estimation benchmarks before generalized architectures can replace specialized pipelines.

References

1. Space Training and Readiness Command (STARCOM). *SDP 3-100, Space Domain Awareness*; United States Space Force Space Doctrine Publication: Colorado Springs, CO, USA 2023.
2. GregorDS. Six Degrees of Freedom—Wikipedia, the Free Encyclopedia. 2015 Available online: https://en.wikipedia.org/wiki/Six_degrees_of_freedom#/media/File:6DOF.svg (accessed on 25 September 2024).
3. Wood, G.E. Estimation of Satellite Orientation from Space Surveillance Imagery Measured with an Adaptive Optics Telescope. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, OH, USA, 1996.
4. Lucas, J.; Kyono, T.; Werth, M.; Gagnier, N.; Endsley, Z.; Fletcher, J.; AFRL, I.M. Estimating Satellite Orientation through Turbulence with Deep Learning. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 15–18 September 2020.
5. Wang, G.; Manhardt, F.; Tombari, F.; Ji, X. GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 16611–16621.
6. Li, Z.; Wang, G.; Ji, X. CDPN: Coordinates-Based Disentangled Pose Network for Real-Time RGB-Based 6-DoF Object Pose Estimation. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7677–7686. [\[CrossRef\]](#)
7. Liu, X.; Zhang, R.; Zhang, C.; Fu, B.; Tang, J.; Liang, X.; Tang, J.; Cheng, X.; Zhang, Y.; Wang, G.; et al. GDRNPP. 2022. Available online: https://github.com/shanice-1/gdrnpp_bop2022 (accessed on 20 September 2024).
8. Zhang, R.; Huang, Z.; Wang, G.; Liu, X.; Zhang, C.; Ji, X. GPose2023: A Modularized Learning-based Object Pose Estimator. In Proceedings of the 8th International Workshop on Recovering 6D Object Pose (ICCV), Paris, France, 3 October 2023.
9. Sundermeyer, M.; Hodan, T.; Labbé, Y.; Wang, G.; Brachmann, E.; Drost, B.; Rother, C.; Matas, J. BOP Challenge 2022 Results Presentation. In Proceedings of the 7th International Workshop on Recovering 6D Object Pose, ECCV 2022, Tel Aviv, Israel, 23 October 2022. Available online: https://cmp.felk.cvut.cz/sixd/workshop_2022/slides/bop_challenge_2022_results.pdf (accessed on 20 October 2024).
10. Hodan, T.; Sundermeyer, M.; Labbe, Y.; Nguyen, V.N.; Wang, G.; Brachmann, E.; Drost, B.; Lepetit, V.; Rother, C.; Matas, J. BOP Challenge 2023 on Detection Segmentation and Pose Estimation of Seen and Unseen Rigid Objects. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 5610–5619.
11. Zhou, Y.; Barnes, C.; Lu, J.; Research, A.; Yang, J.; Li, H. On the continuity of rotation representations in neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
12. Kisantal, M.; Sharma, S.; Park, T.H.; Izzo, D.; Mörtens, M.; D'Amico, S. Satellite pose estimation challenge: Dataset, competition design, and results. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56*, 4083–4098. [\[CrossRef\]](#)
13. Park, T.H.; Mörtens, M.; Jawaid, M.; Wang, Z.; Chen, B.; Chin, T.J.; Izzo, D.; D'Amico, S. Satellite Pose Estimation Competition 2021: Results and Analyses. *Acta Astronaut.* **2023**, *204*, 640–665. [\[CrossRef\]](#)
14. Park, T.H.; D'Amico, S. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. *Adv. Space Res.* **2024**, *73*, 5726–5740. [\[CrossRef\]](#)
15. Park, T.H.; D'Amico, S. Adaptive neural-network-based unscented kalman filter for robust pose tracking of noncooperative spacecraft. *J. Guid. Control Dyn.* **2023**, *46*, 1671–1688. [\[CrossRef\]](#)
16. Park, T.H.; D'Amico, S. Bridging Domain Gap for Flight-Ready Spaceborne Vision. *arXiv* **2024**, arXiv:2409.11661.
17. Dickinson, T.; Walvoord, D.; Gartley, M. Automated 6DOF Satellite Pose Estimation from Resolved Ground-Based Imagery. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 17–20 September 2024.
18. Dickinson, T. From Sim to 6DOF: Deep Learning for Real-Time Satellite Pose Estimation from Resolved Ground-Based Imagery. Ph.D. Thesis, Rochester Institute of Technology, RIT—Main Campus, Henrietta, NY, USA, 2025.
19. Drummond, J.D. Adaptive optics Lorentzian point spread function. In *Adaptive Optical System Technologies*; SPIE: Bellingham, WA, USA, 1998; Volume 3353, pp. 1030–1037.
20. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: <https://www.tensorflow.org> (accessed on 10 August 2025).
21. Astropy Collaboration; Price-Whelan, A.M.; Lim, P.L.; Earl, N.; Starkman, N.; Bradley, L.; Shupe, D.L.; Patil, A.A.; Corrales, L.; Brasseur, C.E.; et al. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *Astrophys. J.* **2022**, *935*, 167. [\[CrossRef\]](#)
22. Adriano, A.; Scott, K.A.; Azad, N.L. Extreme Gradient Boosting and Deep Learning Models for the Classification of Synthetic Space Debris Light Curves. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 17–20 September 2024.

23. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
24. 3rd Generation Partnership Project (3GPP). *Technical Specification Group Radio Access Network; Study on New Radio (NR) to Support Non-Terrestrial Networks (Release 15)*; Technical Report 38.811; 3rd Generation Partnership Project (3GPP): Sophia Antipolis, France, 2019.
25. Ghiasi, G.; Lee, H.; Kudlur, M.; Dumoulin, V.; Shlens, J. Exploring the structure of a real-time, arbitrary neural artistic stylization network. In Proceedings of the British Machine Vision Conference (BMVC), London, UK, 4–7 September 2017.
26. Google. Arbitrary Image Stylization v1. 2020. Available online: <https://www.kaggle.com/models/google/arbitrary-image-stylization-v1/tensorFlow1/256/2?tfhub-redirect=true> (accessed on 1 December 2024).
27. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and flexible image augmentations. *Information* **2020**, *11*, 125. [CrossRef]
28. Goodenough, A.A.; Brown, S.D. DIRSIG5: Next-generation remote sensing data and image simulation framework. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4818–4833. [CrossRef]
29. Por, E. Adaptive Optics with a Shack-Hartmann Wavefront Sensor, 2017–2022. Available online: <https://docs.hcipy.org/dev/tutorials/ShackHartmannWFS/ShackHartmannWFS.html> (accessed on 19 November 2024).
30. Por, E.H.; Haffert, S.Y.; Radhakrishnan, V.M.; Doelman, D.S.; van Kooten, M.; Bos, S.P. High Contrast Imaging for Python (HCIPy): An open-source adaptive optics and coronagraph simulator. In *Adaptive Optics Systems VI*; SPIE: Bellingham, WA, USA, 2018; Volume 10703, pp. 1112–1125.
31. Johnson, R.; Montera, D.; Schneeberger, T.; Spinhirne, J. A new sodium guidestar adaptive optics system for the Starfire Optical Range 3.5 m telescope. In *Adaptive Optics: Methods, Analysis and Applications*; Optica Publishing Group: Washington, DC, USA, 2009; p. AOTuA1.
32. Gladysz, S.; Le Louarn, M.; Yaitskova, N.; Garcia-Rissmann, A.; Kann, L.; Drummond, J.D.; Johnson, R.L.; Roskey, D. Size of the halo of the adaptive optics PSF. In *Adaptive Optics Systems III*; SPIE: Bellingham, WA, USA, 2012; Volume 8447, pp. 802–816.
33. Fried, D.L. Optical resolution through a randomly inhomogeneous medium for very long and very short exposures. *J. Opt. Soc. Am.* **1966**, *56*, 1372–1379. [CrossRef]
34. Tyson, R.K.; Frazier, B.W. *Principles of Adaptive Optics*; CRC Press: Boca Raton, FL, USA, 2022.
35. Zheng, Z.; Wang, P.; Ren, D.; Liu, W.; Ye, R.; Hu, Q.; Zuo, W. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Trans. Cybern.* **2021**, *52*, 8574–8586. [CrossRef] [PubMed]
36. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 12993–13000.
37. Ren, S. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv* **2015**, arXiv:1506.01497.
38. Kingma, D.P. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
39. Kalman, R.E. A new approach to linear filtering and prediction problems. *ASME J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]
40. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*; John Wiley & Sons: Hoboken, NJ, USA, 2001.
41. SciPy Developers SciPy: Scientific Computing Tools for Python—Rotation.Magnitude. 2024. Available online: <https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.transform.Rotation.magnitude.html> (accessed on 5 October 2024).
42. NASA. Hubble Space Telescope 3D Model. NASA Science Resource. 2019. Available online: <https://science.nasa.gov/resource/hubble-space-telescope-3d-model-2/> (accessed on 14 August 2025).
43. Wang, G.; Manhardt, F.; Tombari, F.; Ji, X. GDR-Net: Geometry-Guided Direct Regression Network for Monocular 6D Object Pose Estimation. GitHub Repository. 2021. Available online: <https://github.com/THU-DA-6D-Pose-Group/GDR-Net> (accessed on 14 August 2025).
44. Rochester Institute of Technology. Research Computing Services. Available online: https://research-computing.git-pages.rit.edu/docs/about_us.html (accessed on 1 July 2023).
45. Montera, D. Passing the Torch: Creating Our Own Stars. YouTube Video, AFRL Inspire, 2017. Relevant Data Displayed at 11m30s. Available online: <https://www.youtube.com/watch?v=s6pE2GxJGm4> (accessed on 1 July 2023).
46. Air Force Research Laboratory. High-Resolution Satellite Imagery from the SOR 3.5 m Telescope. 1996. Available online: <https://spp.fas.org/military/program/track/starfire.pdf> (accessed on 1 December 2024).
47. Werth, M.; Brandoch, C.; Roe, K.; Conti, A. Multi-Frame Blind Deconvolution Accelerated with Graphical Processing Units. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 17–20 September 2019.
48. Fulcoly, D.O.; Kalamaroff, K.I.; Chun, F.K. Determining basic satellite shape from photometric light curves. *J. Spacecr. Rocket.* **2012**, *49*, 76–82. [CrossRef]

49. Bennett, D.; Allen, D.; Dank, J.; Gartley, M.; Tyler, D. SSA Modeling and Simulation with DIRSIG. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 9–12 September 2014.
50. Leachtenauer, J.C.; Malila, W.; Irvine, J.; Colburn, L.; Salvaggio, N. General image-quality equation: GIQE. *Appl. Opt.* **1997**, *36*, 8322–8328. [[CrossRef](#)] [[PubMed](#)]
51. Thurman, S.T.; Fienup, J.R. Analysis of the general image quality equation. In *Visual Information Processing XVII*; SPIE: Bellingham, WA, USA, 2008; Volume 6978, pp. 102–114.
52. Lucas, J.; Kyono, T.; Yang, J.; Fletcher, J. Discovering 3-D Structure of LEO Objects. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 14–17 September 2021.
53. Hodaň, T.; Matas, J.; Sundermeyer, M.; Huang, J.; Fourmy, M.; Nguyen, V.N.; Kalra, A.; Taamazyan, V.; Tran, C.; Tyree, S.; et al. Benchmark for 6D Object Pose Estimation (BOP Challenge). 2025. Available online: <https://bop.felk.cvut.cz/challenges/> (accessed on 20 August 2025).
54. Legrand, A.; Detry, R.; De Vleeschouwer, C. Domain Generalization for In-Orbit 6D Pose Estimation. *arXiv* **2024**, arXiv:2406.11743.
55. Jiang, X.; Missel, R.; Li, Z.; Wang, L. Sequential Latent Variable Models for Few-Shot High-Dimensional Time-Series Forecasting. In Proceedings of the 2023 International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023.
56. Ye, Y.; Toloubidokhti, M.; Vadhavkar, S.; Jiang, X.; Liu, H.; Wang, L. On the Identifiability of Hybrid Deep Generative Models: Meta-Learning as a Solution. In *Advances in Neural Information Processing Systems*; Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2024; Volume 37, pp. 7714–7735.
57. Badura, G.P.; Velez-Reyes, M.; Gunter, B.; Valenta, C.R.; Ho, K. Regularizing Training of Physics Informed Neural Networks (PINNs) for Cislunar Orbit Determination via Transfer Learning. In Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference (AMOS), Maui, HI, USA, 17–20 September 2024.
58. Ulmer, M.; Durner, M.; Sundermeyer, M.; Stoiber, M.; Triebel, R. 6d object pose estimation from approximate 3d models for orbital robotics. In *Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2023; pp. 10749–10756.
59. Zhong, L.; Chen, S.; Wang, W.; Yang, W.; Yuan, G.; Guo, P.; Yang, X.; Zhang, X. Uncooperative Spacecraft Pose Estimation with Normalized Segmentation Coordinate Space. *IEEE/ASME Trans. Mechatron.* **2024**, *30*, 2293–2304. [[CrossRef](#)]
60. OpenAI. Introducing 4o Image Generation. 2025. Available online: <https://openai.com/index/introducing-4o-image-generation/> (accessed on 22 April 2025).
61. OpenAI. Introducing OpenAI o3 and o4-Mini. 2025. Available online: <https://openai.com/index/introducing-o3-and-o4-mini/> (accessed on 22 April 2025).
62. OpenAI. Thinking with Images. 2025. Available online: <https://openai.com/index/thinking-with-images/> (accessed on 22 April 2025).
63. Google. Gemini 3.1 Pro. 2026. Available online: <https://blog.google/innovation-and-ai/models-and-research/gemini-models/gemini-3-1-pro/> (accessed on 19 March 2026).
64. Roboflow. Vision Evals: Visual Understanding. 2026. Available online: <https://playground.roboflow.com/evals?category=visual-understanding&task=Spatial+Understanding> (accessed on 26 May 2026).
65. Yang, L.; Kang, B.; Huang, Z.; Zhao, Z.; Xu, X.; Feng, J.; Zhao, H. Depth anything v2. *Adv. Neural Inf. Process. Syst.* **2024**, *37*, 21875–21911. [[CrossRef](#)]
66. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2009; pp. 248–255.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.