AAS 26-962

RECONSTRUCTION-DRIVEN REINFORCEMENT LEARNING FOR AUTONOMOUS SATELLITE INSPECTION WITH GAUSSIAN SPLATTING

Logan Selph* **and John R. Martin†**

Recent work in reinforcement learning policies for spacecraft inspection has focused on optimizing coverage of a discretized spherical approximation of the target, even though the desired inspection product is an accurate geometric model. This separation between the planning objective and the reconstruction product can reward observations that add little information while excluding views that remain useful for recovering poorly reconstructed geometry. We investigate whether coupling reinforcement learning with an active active reconstruction method can overcome these challenges by incorporating a global geometry approximation into the optimization process. We propose Splat-RL, a reinforcement-learning framework that derives its observation and image-capture reward from the reconstruction error of a Gaussian Splatting model. We compare Splat-RL with an illumination-aware coverage policy using this spherical heuristic. We find that the coverage policy fails to discover full geometric detail because its spherical spacecraft approximation neglects complex shadowing effects, whereas Splat-RL selects high-illumination and informative views that reconstruct the missing structure and improve all reported three-dimensional reconstruction metrics in our experiments. These results show that the accurate photometric information produced by a GS model can be combined with RL approaches to reliably find better inspection trajectories compared to current coverage based spherical heuristic models.

INTRODUCTION

The increasing number of spacecraft in low Earth orbit (LEO) has strengthened the need for reliable and autonomous rendezvous and proximity operations (RPO) capabilities. Typically, individual RPO-focused missions require a dedicated ground team to ensure that all maneuvers are safe and optimal, since proximity operations are delicate tasks that risk the health of both the target and chaser. Organizing these missions at a larger scale means that we must significantly advance our ability to carry out any, or all, of these required tasks without the explicit aid of a dedicated ground team. A significant portion of RPO-focused missions involves inspecting the target resident space object (RSO) so that later decisions can account for its geometry, condition, and operational state. An autonomous inspector could reduce this operational burden by selecting viewpoints, setting trajectories, and collecting images with little to no ground intervention. To satisfy this, the resulting planning method should optimize both the quality of the resulting inspection product as well as the path followed by the chaser.

Existing autonomous inspection formulations often simplify the problem by approximating the RSO as a sphere. A spherical approximation in this case accomplishes two things: illumination estimates become trivial to calculate, and the RSO's surface can easily be discretized so the inspection

*NSF Graduate Research Fellow, Department of Aerospace Engineering, University of Maryland, College Park, MD.
†Assistant Professor, Department of Aerospace Engineering, University of Maryland, College Park, MD.

task can be posed as a maximum coverage problem. This provides a tractable and target-agnostic objective, but spherical coverage does not guarantee that the collected images will improve a reconstruction of the true underlying geometry. The solutions provided by motion planning policies for this environment can consistently provide excellent spacecraft coverage, but the drastic oversimplification of complex shadowing effects means that the captured images could often provide little useful information. Modern machine learning based rendering methods provide a direct way to simultaneously model the geometric and photometric structure of a scene, providing fast feedback about next-best-view choices. This work uses that capability to construct a reinforcement-learning objective from the error of a gaussian splatting (GS) model to investigate whether reconstruction-aware planning produces more useful inspection images than an illumination-aware coverage heuristic.

Our work makes three primary contributions. First, it formulates spacecraft inspection around the measured error of the reconstruction product instead of a geometry-independent spherical coverage proxy. Second, it supplies a PPO policy with a compact global representation of the directional PSNR error by predicting spherical-harmonic coefficients from the body-frame Sun direction. Third, it compares the resulting Splat-RL policy with a reproduction of the spherical heuristic coverage formulation under the same dynamics, initialization domain, and 31-image budget. The experiment intentionally removes training views that support a single solar panel, creating a controlled reconstruction defect that both policies must address. The resulting comparison reveals how the baseline illumination rule neglects important shadowing effects, while Splat-RL does not.

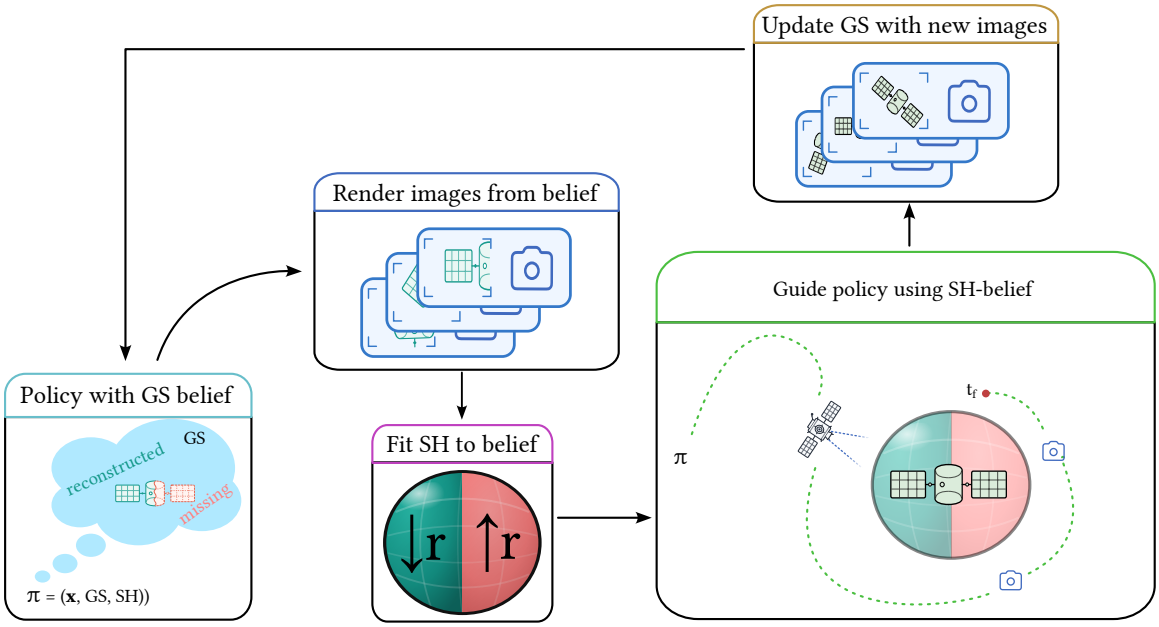


Figure 1: Overview of the proposed Splat-RL pipeline. The framework evaluates a coarse GS reconstruction, compresses its directional error into a Sun-conditioned spherical-harmonic representation, supplies that representation to the policy, and uses the selected images to refine the reconstruction.

BACKGROUND AND RELATED WORK

Spacecraft Inspection Trajectory Planning

Autonomous spacecraft inspection has traditionally been formulated as a motion-planning problem in which relative dynamics are used to construct trajectories that balance observation goals against safety, mission duration, and fuel use. Optimization-based methods have demonstrated that proximity trajectories can be generated while explicitly enforcing collision-avoidance and other operational constraints [1]. Sampling-based approaches extend this framework by efficiently searching constrained state spaces under orbital dynamics, enabling propellant-efficient proximity operations and inspection-specific planning [2, 3]. These methods have been applied to full-coverage inspection, coordinated multi-spacecraft operations, receding-horizon planning around tumbling or non-cooperative targets, and passive-safety constraints [4, 5, 6]. Information-driven methods instead select trajectories according to the expected value of future observations, allowing sensing objectives to be balanced directly against maneuver and energy costs [7]. Related formulations incorporate formal safety requirements, coordinated motion, and Sun-avoidance constraints directly into the planning architecture [8]. Collectively, these model-based approaches provide interpretable trajectories and explicit treatment of mission constraints, but changes in geometry, illumination, spacecraft state, or sensing objectives generally require a new trajectory search or optimization.

Learning-based methods instead seek to refine part of this decision-making process by learning a policy that maps the current spacecraft and inspection state directly to maneuver or observation decisions. Early deep-reinforcement-learning approaches demonstrated learned guidance for spacecraft proximity operations and rendezvous [9, 10]. Subsequent work treated inspection itself as the decision-making objective, incorporating illumination and target coverage into learned policies [11]. More recent formulations introduce variable-duration actions, optimization-based safety shields, and runtime assurance for higher-fidelity spacecraft dynamics and operational constraints [12, 13]. Learning-based inspection has also been extended toward uncertain target reconstruction and decentralized multi-agent viewpoint selection [14, 15]. Despite these advances, most existing inspection policies still define sensing value primarily through geometric coverage or related surrogate objectives rather than through the quality of the reconstructed target model. Splat-RL addresses this distinction by deriving the inspection reward directly from the current reconstruction error, allowing the policy to prioritize observations according to their expected value based on the evolving scene representation.

Active View Planning for Reconstruction

Reconstruction-driven sensing is commonly formulated as active vision or next-best-view (NBV) planning, in which camera poses are selected according to their expected ability to reduce uncertainty, reveal unobserved structure, or improve reconstruction quality. Neural scene representations have enabled uncertainty estimated directly from the current reconstruction to guide subsequent image acquisition [16]. Reinforcement learning (RL) has similarly been used to learn generalizable NBV policies over continuous viewpoint spaces [17]. These approaches establish a baseline which we can adapt into new NBV formulations that apply directly to orbital environments. Spacecraft inspection introduces additional structure not typically present in terrestrial NBV planning. The nature of the governing dynamics and state uncertainty changes significantly, fuel constraints drive the need for impulsive maneuvers with variable-time coasting periods, and illumination conditions become much more restrictive based on lack of diffuse lighting and complex shadows. A

reconstruction-driven spacecraft policy must therefore couple NBV selection with astrodynamics and illumination-aware sensing. Splat-RL addresses this problem by encoding reconstruction error as a function of viewing state, based on the spacecraft position and the solar illumination direction.

Three-Dimensional Spacecraft Reconstruction

Three-dimensional spacecraft reconstruction spans traditional geometric and photometric techniques as well as learned radiance-field and point-based representations. Structure-from-motion estimates camera pose and scene geometry from feature correspondences across multiple images and remains a standard approach for image-based reconstruction [18]. Spacecraft imagery can challenge feature-based methods because of reflective materials, repeated structure, low-texture regions, and strongly varying illumination. Photometric techniques such as stereophotoclinometry combine stereo geometry with shape-from-shading information to recover detailed surface structure and have been demonstrated extensively in space-imaging applications [19]. Newer techniques based on stereophotoclinometry have further adapted it to require less human-in-the-loop assistance, and bolster its rendering quality [20].

Recently, machine learning based rendering techniques have gained attention from the scientific community due to their prowess at solving the inverse rendering problem quickly with less information than traditional techniques. Neural Radiance Fields represent scenes as continuous functions of spatial position and viewing direction and use differentiable volumetric rendering to synthesize novel observations [21]. Three-dimensional Gaussian Splatting instead represents the scene explicitly using anisotropic gaussian primitives, enabling substantially faster optimization and rendering while maintaining high-quality novel-view synthesis [22]. Recent spacecraft-specific work has demonstrated accelerated Gaussian Splatting for reconstruction of unknown satellite geometry and investigated its suitability for spaceflight-relevant computational environments [23]. Spacecraft-focused GS has also been extended to incorporate known Sun-direction information, reducing ambiguity introduced by changing illumination and shadowing conditions [24]. These properties make Gaussian Splatting particularly attractive for reconstruction-in-the-loop inspection, where the scene representation must be repeatedly updated and evaluated during policy execution.

PRELIMINARIES

Gaussian Splatting

A GS model represents a scene with a set of anisotropic three-dimensional Gaussian primitives. Each primitive contains a center $\boldsymbol{\mu}_i$, a covariance $\boldsymbol{\Sigma}_i$, an opacity parameter, and appearance coefficients that can vary with viewing direction. Implementations commonly factor the covariance into a rotation and axis-aligned scale so that optimization preserves a valid positive-semidefinite shape. The renderer projects each three-dimensional Gaussian into the image plane, orders the projected primitives by depth, and blends their contributions at every affected pixel. This explicit representation allows a tiled rasterizer to skip primitives that do not influence a pixel and gives GS its primary speed advantage over ray-marched neural fields [22].

For a pixel coordinate $\mathbf{u}$, the renderer combines the depth-ordered primitive colors through alpha compositing,

$$\mathbf{C}(\mathbf{u}) = \sum_{i=1}^{N} T_i(\mathbf{u})\, \hat{\alpha}_i(\mathbf{u})\, \mathbf{c}_i, \tag{1}$$

where $T_i(\mathbf{u})$ denotes the accumulated transmittance before primitive i, $\hat{\alpha}_i(\mathbf{u})$ denotes its projected opacity contribution, and $\mathbf{c}_i$ denotes its color for the requested viewing condition. Training renders images from known camera poses and compares those renders with corresponding ground-truth images. The optimizer differentiates the photometric loss through the rasterizer and updates primitive position, shape, opacity, and appearance. A basic squared photometric objective takes the form

$$\mathcal{L}_{\text{rgb}} = \sum_{\mathbf{u}} \|\mathbf{C}(\mathbf{u}) - \mathbf{C}^*(\mathbf{u})\|_2^2, \tag{2}$$

where $\mathbf{C}^*(\mathbf{u})$ denotes the reference pixel value. Practical GS training also densifies under-resolved regions and prunes primitives that contribute little, which lets the representation allocate capacity where the observed images demand it.

Spherical-Harmonics

To represent the illumination-aware construction error-based reward when training our RL agent, we have opted to use a single lightweight spherical harmonics model. Spherical harmonics provide an orthogonal basis for scalar functions defined over a sphere. A real spherical-harmonic expansion represents a directional function by weighting basis functions indexed by degree ℓ and order m. For a unit viewing direction $\mathbf{v}$ with azimuth ϕ and polar angle θ, the complex spherical-harmonic basis is

$$Y_\ell^m(\theta, \phi) = N_{\ell m} P_\ell^m(\cos\theta) e^{im\phi}, \tag{3}$$

where P_ℓ^m is the associated Legendre polynomial and $N_{\ell m}$ is the normalization constant. We use the corresponding real spherical-harmonic basis,

$$Y_{\ell m}(\theta, \phi) = \begin{cases} \sqrt{2}(-1)^m \operatorname{Im}\left(Y_\ell^{|m|}(\theta, \phi)\right), & m < 0, \\ Y_\ell^0(\theta, \phi), & m = 0, \\ \sqrt{2}(-1)^m \operatorname{Re}\left(Y_\ell^m(\theta, \phi)\right), & m > 0. \end{cases} \tag{4}$$

This work uses real spherical harmonics through degree $L = 4$, giving $(L+1)^2 = 25$ basis functions.

Typically, a spherical harmonics model will achieve its fit by regressing a single coefficient vector of dimension equal to the number of basis functions. This works well for static distributions that can be represented over a single sphere, but does not extend naturally if the same distribution also has other time related dependencies. In the case of in-situ spacecraft modeling, the quality of reconstructed images depends both on the requested viewing direction and the time-dependent illumination condition. Some natural choices to account for this Sun-angle information include tensor-product spherical harmonics expansions and interpolation. Tensor-product spherical harmonic expansions approach the issue by adding an additional set of basis functions for the extra dependency and regressing a much larger coefficient matrix [25]. While this retains the same efficient linear regression methods as before, the new coefficient matrix inflates quickly. Alternatively, it is possible to generate a grid of SH models for different lighting conditions and interpolate between them. However, a sparse sampling of illumination conditions can lose structure between sampled lighting directions, while accurately representing more complex illumination-dependent behavior requires a much denser sampling of the illumination domain [26]. This substantially increases the number of models that must be generated and stored. Instead, we have chosen to teach a neural network

to predict the final coefficient vector based on the current Sun illumination condition. Let $\mathbf{s} \in \mathbb{R}^3$ denote the body-frame Sun unit vector and let $\mathbf{v} \in \mathbb{R}^3$ denote the body-fixed unit vector from the RSO to the chaser. A multilayer perceptron f_θ maps the Sun direction to the coefficient vector,

$$\mathbf{c}(\mathbf{s}) = f_\theta(\mathbf{s}). \tag{5}$$

If $\mathbf{y}(\mathbf{v}) \in \mathbb{R}^{25}$ contains the real spherical-harmonic basis evaluated at $\mathbf{v}$, the model predicts the normalized reconstruction-error reward as

$$\hat{r}_{\mathrm{GS}}(\mathbf{s}, \mathbf{v}) = \mathrm{clip}\left(\mathbf{y}(\mathbf{v})^\top \mathbf{c}(\mathbf{s}), 0, 1\right). \tag{6}$$

The coefficient network receives the three Cartesian components of $\mathbf{s}$, uses three hidden layers of width 64 with SiLU activations, and outputs 25 coefficients. Offline training minimizes mean-squared error with AdamW at a learning rate of 3×10^{-3}, a weight decay of 10^{-5}, and 5000 epochs.

Proximal Policy Optimization

Proximal Policy Optimization (PPO) trains an actor–critic policy from trajectories collected through interaction with the environment [27]. The actor parameterizes the action distribution $\pi_\phi(\mathbf{a}_t \mid \mathbf{o}_t)$, while the critic estimates the value of the current observation. PPO reuses each rollout for several minibatch updates but limits policy changes through a clipped surrogate objective. If $\rho_t(\phi)$ denotes the probability ratio between the updated and rollout policies and $\hat{A}_t$ denotes the estimated advantage, the clipped objective includes

$$\mathcal{L}_{\mathrm{clip}}(\phi) = \mathbb{E}_t\left[\min\left(\rho_t(\phi)\hat{A}_t, \mathrm{clip}(\rho_t(\phi), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]. \tag{7}$$

This clipping discourages a single update from moving the policy too far from the behavior that generated the data. Stable-Baselines3 supplies the base PPO implementation, and the present work modifies its rollout and return calculations for variable-duration decisions.

The inspection problem uses a semi-Markov decision process (sMDP) because each action includes a coast duration τ_t. The rollout buffer stores that duration and applies a transition-dependent discount,

$$\gamma_t = \exp(-\beta\tau_t), \qquad \beta = 10^{-4}. \tag{8}$$

The generalized advantage and return calculations use γ_t rather than a single fixed discount for every decision. The policy evaluates an image request at the beginning of a decision, applies the impulsive maneuver, and then propagates the state for the selected coast interval. This event-based structure matches the inspection task because the policy receives reconstruction reward when it captures an image rather than continuously throughout the coast.

SPLAT-RL FRAMEWORK

Overview

In this work we demonstrate Splat-RL's performance on synthetic datasets based on an ACRIMSAT model provided by the NASA 3D model archive. Our framework uses PyTorch3D to generate ground-truth spacecraft images, a modified Nerfstudio repository to manage GS models [28], Gymnasium to implement the inspection environment, and the modified Stable-Baselines3 PPO implementation to train the policy. Figure 1 summarizes the the structure of Splat-RL, notably

highlighting that the gradients between the GS and PPO models are separated. Our focus is to investigate how policies are influenced by reward landscapes generated from trained GS models, so reconstruction metrics only influence observation rewards collected by the PPO agent. To ensure consistent and repeatable analysis of this influence, our experiments are initialized with a crude pretrained model rather than starting the inspection process from scratch. By starting with this prior information, we can investigate how our framework impacts inspection trajectories and observation behavior, and if it can iteratively improve the final reconstruction as the GS model is updated.

PSNR-Derived Reward

Sparse rewards are based on the reconstruction error associated with the position and lighting condition of a requested observation. We have chosen PSNR, a log scaled error metric, to represent the performance of our reconstructions,

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{1}{\mathrm{MSE}} \right). \tag{9}$$

where pixel values are normalized and MSE represents the mean squared error between the ground truth and predicted images. A low PSNR indicates that the model reproduces the reference image poorly, identifying a potentially informative observation. The framework maps the PSNR, p, to a reward according to

$$r_{\mathrm{PSNR}}(p) = \mathrm{clip} \left(\frac{35 - p}{35 - 20}, 0, 1 \right). \tag{10}$$

Views at or above 35 dB receive zero target reward, values between 20 and 35 dB map linearly to $[0, 1]$, and values at or below 20 dB receive the maximum target reward. The current Splat-RL formulation sets the reward to zero during eclipse but does not impose the coverage baseline's additional requirement that a surface waypoint lie within 60° of the Sun direction. Equation 10 measures current reconstruction error compared to the GS model, rather than the exact marginal improvement that could be caused by a new image. Our experiment tests whether this error remains a useful and tractable surrogate for NBV selection.

The initial reward-model dataset contains 24 Sun contexts distributed around the circular target orbit, as Figure 2 shows. The contexts differ by approximately 15° in body-frame azimuth and have zero body-frame elevation. For each context, the framework evaluates the incomplete GS model at 200 full-sphere viewing directions, which produces 4800 PSNR and reward samples. It holds out Sun-context indices $\{0, 6, 12, 18\}$, leaving 4000 samples from 20 contexts for fitting and 800 samples from four contexts for validation. Heatmap interpolation supports visualization only, while the spherical-harmonic model trains directly on the sampled viewing directions and reward targets.

SMDP Formulation

The inspection problem is represented by the semi-Markov decision process

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, P, R, \mathcal{T}, \Gamma), \tag{11}$$

where $\mathcal{S}$ is the latent state space, $\mathcal{A}$ is the action space, P is the variable-duration transition kernel, R is the event-based reward function, $\mathcal{T}$ is the holding-time model, and Γ is the duration-dependent discount model.

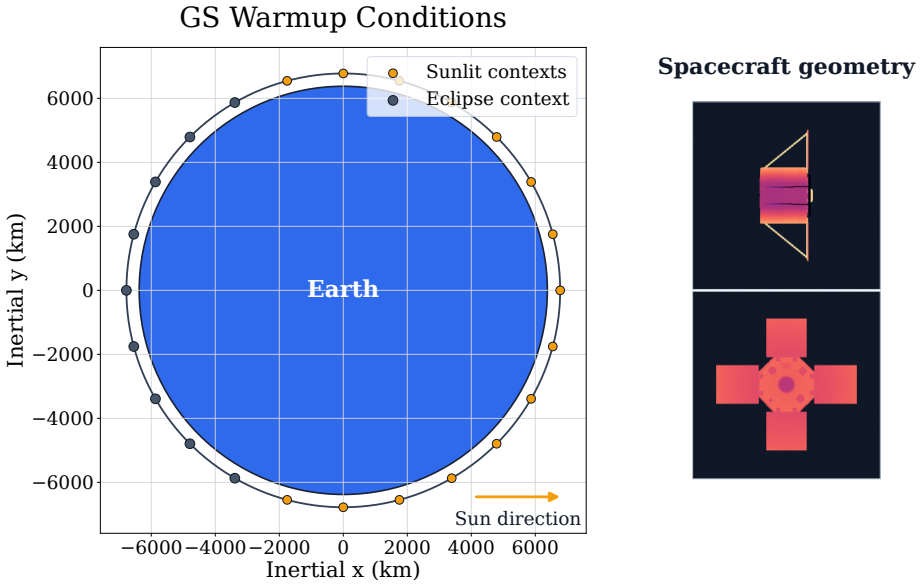


Figure 2: Orbital sample locations used to generate the initial coarse-GS and reconstruction-error datasets. The 24 true-anomaly conditions include both illuminated and eclipse portions of the circular RSO orbit.

The Splat-RL observation contains the relative translational state, orbital phase, illumination information, capture-budget information, capture-validity indicators, and the compressed GS error landscape. The implementation represents the observation as

$$\mathbf{o}_t = [\mathbf{x}_t, \dot{\mathbf{x}}_t, b_t, \nu_t, \mathbf{s}_t, I_{\text{valid},t}, I_{\text{novel},t}, \hat{\mathbf{c}}(\mathbf{s}_t)] \,, \tag{12}$$

where $\mathbf{x}_t$ and $\dot{\mathbf{x}}_t$ denote the chaser position and velocity in the target LVLH frame, b_t denotes the consumed fraction of the image budget, ν_t denotes RSO's true anomaly, and $\mathbf{s}_t$ denotes the body-frame Sun direction. The validity indicator identifies whether a capture satisfies range and eclipse conditions, while the novelty indicator identifies whether the current direction differs sufficiently from prior successful captures. The coefficient vector gives the policy a global description of the current directional objective, but it does not encode the complete capture history. This combination lets the policy relate its current orbital state to the full reconstruction-error landscape at each decision.

Each policy decision produces the continuous action

$$\mathbf{a}_t = [\Delta v_x, \Delta v_y, \Delta v_z, u_\tau, u_{\text{cap}}] \,. \tag{13}$$

The first three components define an impulsive LVLH velocity change with a maximum norm of 0.1 m/s per decision. The duration command $u_\tau \in [-1, 1]$ maps to a coast duration between 10 and 300 s, and the capture command requests an image when it exceeds the decision threshold. The environment evaluates a requested image at the current state before it applies the maneuver and propagates the coast. The experiment imposes neither a cumulative ΔV limit nor an explicit ΔV penalty because it seeks to isolate the consequence of the sensing objective rather than fuel-optimal behavior.

A candidate capture counts as novel when its body-frame viewing direction lies at least 10° from every direction in the active capture buffer. Both policies receive a budget of 31 successful imaging opportunities, so the active buffer can represent the complete set of successful captures in the current rollout. Splat-RL grants a positive image reward only when the policy requests a capture and the view satisfies the geometric, eclipse, and novelty conditions,

$$r_{\mathrm{img},t} = I_{\mathrm{cap},t} I_{\mathrm{valid},t} I_{\mathrm{novel},t} \hat{r}_{\mathrm{GS}}(\mathbf{s}_t, \mathbf{v}_t). \tag{14}$$

The geometric and eclipse validity indicator takes the form

$$I_{\mathrm{valid},t} = \mathbb{I}[r_{\mathrm{coll}} < \|\mathbf{x}_t\| \leq r_{\mathrm{insp}}]\, \mathbb{I}[\neg\, \mathrm{eclipse}(t)]\,. \tag{15}$$

EXPERIMENTAL SETUP

Orbital and Inspection Environment

The target ACRIMSAT model follows a circular orbit at an altitude of 400 km with zero inclination and zero eccentricity. The target remains nadir locked, so its body frame rotates with the local-vertical local-horizontal frame throughout the orbit. The simulation treats the Sun direction as fixed in the inertial frame, which causes the body-frame Sun direction to change with true anomaly. The environment explicitly models Earth eclipse and prevents Splat-RL from receiving a positive image reward during eclipse. The chaser camera always points toward the target center, and the current study omits independent target and chaser attitude dynamics.

The chaser state uses an RSO-centered LVLH frame, and the simulator propagates relative translation with the Hill–Clohessy–Wiltshire equations. Each episode samples the initial chaser position within a spherical shell extending from 8 to 250 m and samples the initial relative velocity from the implemented training distribution. Randomized target true anomaly changes the initial illumination context and eclipse timing across episodes. The collision, inspection, and keep-in radii equal 1, 250, and 1000 m, respectively. An episode terminates when the chaser enters the collision region, exits the keep-in sphere, exhausts its 31-image budget, or reaches the ten-orbit duration limit.

Coarse Gaussian Splatting Model

The experiment begins with the same incomplete GS model for both policies. The initial training set contains 720 PyTorch3D-rendered images distributed across 24 Sun-angle conditions and viewing directions that span the unit sphere. The dataset strategically withholds a viewing region spanning 5/6 of the sphere, centered on one of the spacecrafts solar panels. Removing this region deprives the GS model of the views needed to reconstruct a large portion of ACRIMSAT, including one of its solar panels. This controlled defect creates a direct test of whether either inspection objective returns to the region that supports the missing geometry. The initial GS model trains for 30,000 optimization iterations, with each policy rollout starting from the same crude reconstruction.

Compared Inspection Policies

The first agent implements the illumination-aware coverage heuristic based on the sMDP formulation of Stephenson et al. [12]. The environment uses the coarse-GS training images to determine which spherical coverage waypoints count as already inspected before the policy begins its rollout. The policy observes this initialized coverage state and learns a trajectory that seeks the remaining previously unobserved waypoints. A candidate waypoint must satisfy the baseline's observation

rules, including the requirement that the illuminated surface normal lie within 60° of the Sun direction. The present reproduction omits Stephenson et al.'s safety shield because the comparison targets the inspection objective rather than runtime assurance.

The second agent implements Splat-RL. Its environment trains the Sun-conditioned spherical-harmonic MLP on the PSNR errors of the coarse GS model and supplies the resulting coefficient vector to the policy. The agent receives image reward in directions with high predicted reconstruction error, subject to the range, eclipse, and novelty conditions described in Section . Splat-RL does not inherit the coverage baseline's 60° illumination cutoff, so it can value illuminated near-eclipse views when the GS error map assigns them high reward. Both agents train without a cumulative ΔV constraint or fuel penalty because the experiment focuses on failure modes caused by the sensing objective.

Matched-Budget Rollout and Reconstruction Update

After training, each learned policy performs an evaluation rollout and receives exactly 31 imaging opportunities. The rollout records the trajectory, image-capture positions, body-frame viewing directions, target true anomaly, Sun direction, and pre-capture GS PSNR. Each policy's selected images are added to the coarse-GS training set, and a new GS model is trained from scratch to compare. The refinement process uses matched optimization settings, learning rates, densification behavior, renderer configuration, and iteration counts for the two updated models. Additionally we will inspect the effect of iterating once off an image set taken by our Splat-RL agent. In the case of a coverage-focused agent, extra effort would have to be focused towards evaluating point performance and designing way points for another trajectory to seek. Splat-RL's iteration can follow naturally from an observation campaign. After capturing the first set of images, we can simply update the 3D model and corresponding SH error map, and then send the agent out once more to collect a new set of images based on this new reward map. This new 62-image total update will also be compared against the results of the initial 31-image rollout. Each experiment renders all three models over the same held-out views and extracts geometry with the same truncated signed distance function meshing procedure.

Evaluation Metrics and Representation

The photometric evaluation reports PSNR over a fixed set of held-out camera poses and Sun directions, ensuring a complete picture of the reconstruction accuracy. Directional PSNR heatmaps have been chosen to give a general global view of the drift in photometric performance. These heatmaps are represented in polar coordinates, representing the sphere of possible viewing directions surrounding the RSO. Two cases of these heatmaps are presented: northern and southern hemisphere, and full-sphere. All unlabeled polar plots represent the full-sphere layout, where the north pole is the center and the south pole lies across the circumference.

The geometric evaluation reports Chamfer distance (CD), 95^{th} percentile Hausdorff distance (HD95), and absolute Hausdorff distance (HD). The pipeline extracts meshes from each GS model and compares them in the common simulation frame without artificial alignment. Let $\mathcal{P}$ and $\mathcal{Q}$ denote point sets sampled from the reconstructed and ground-truth surfaces, respectively. The symmetric Chamfer distance is defined as

$$d_{\text{CD}}(\mathcal{P},\mathcal{Q}) = \frac{1}{2}\left[\frac{1}{|\mathcal{P}|}\sum_{\mathbf{p}\in\mathcal{P}}\min_{\mathbf{q}\in\mathcal{Q}}\|\mathbf{p}-\mathbf{q}\|_2 + \frac{1}{|\mathcal{Q}|}\sum_{\mathbf{q}\in\mathcal{Q}}\min_{\mathbf{p}\in\mathcal{P}}\|\mathbf{q}-\mathbf{p}\|_2\right]. \tag{16}$$

Chamfer distance measures the average nearest-surface discrepancy in both directions, making it a useful measure of overall geometric agreement while remaining relatively insensitive to isolated local errors. Hausdorff distance instead characterizes the largest nearest-surface discrepancies. Defining the bidirectional set of nearest-neighbor errors as

$$\mathcal{D}(\mathcal{P}, \mathcal{Q}) = \left\{ \min_{\mathbf{q} \in \mathcal{Q}} \|\mathbf{p} - \mathbf{q}\|_2 \, : \, \mathbf{p} \in \mathcal{P} \right\} \cup \left\{ \min_{\mathbf{p} \in \mathcal{P}} \|\mathbf{q} - \mathbf{p}\|_2 \, : \, \mathbf{q} \in \mathcal{Q} \right\}, \tag{17}$$

the absolute Hausdorff distance is

$$d_{\mathrm{HD}}(\mathcal{P}, \mathcal{Q}) = \max \mathcal{D}(\mathcal{P}, \mathcal{Q}). \tag{18}$$

HD95 is the 95^{th} percentile value of these distances, and provides a robust measure of near-worst-case surface disagreement without allowing a small number of outliers to dominate the metric. Absolute HD captures the single largest geometric discrepancy between the reconstructed and reference surfaces. Chamfer and Hausdorff distances are normalized by the target characteristic diameter D, allowing geometric errors to be compared consistently across targets of different physical scales. Lower Chamfer and Hausdorff distances indicate better surface agreement.

RESULTS

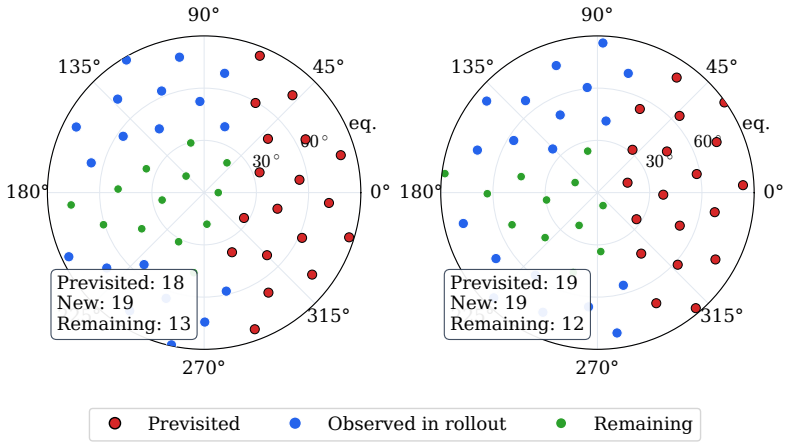
Policy Behavior and Selected Views

The coverage policy continues to pursue broad completion of the unobserved spherical waypoints. Its trajectory distributes image captures across the remaining admissible coverage regions, even when some of those directions already have moderate reconstruction quality. The strict illumination rule removes a substantial portion of the near-eclipse viewing domain from consideration for the nadir-pointing target. As a result, the learned policy does not visit the critical views that reveal the intentionally omitted solar panel. Figure 3 shows the relationship between this broad coverage behavior, the valid collection landscape, and the final 31 capture locations.

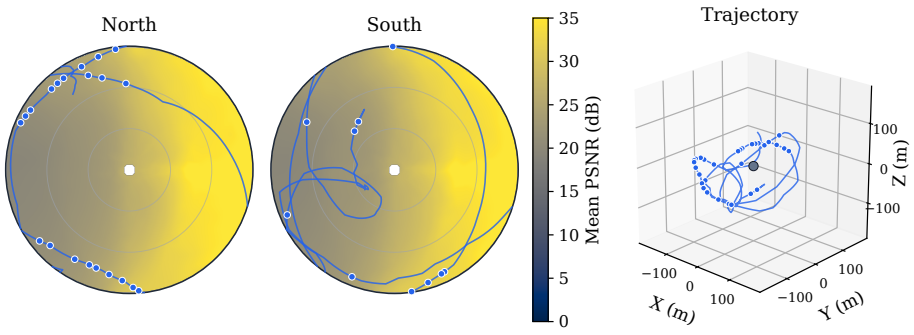
Splat-RL consistently targets the high-error regions of the initial GS model. Since the observable spherical harmonics coefficients give a persistent view of the global reward landscape, NVB selection can consistently target high value viewpoints. Unlike the coverage baseline, Splat-RL can request images from illuminated viewpoints near eclipse even when the angle between a corresponding surface normal and the Sun exceeds 60°. The advantage Splat-RL has over the heuristic policy with respect to image illumination is highlighted in Figure 6. While both Splat-RL and coverage heuristic policies achieve significant spacecraft coverage, Splat-RL consistently captures views with greater illumination quality. Figure 4 shows this concentrated interaction with the reconstruction-error landscape. Additionally, after a 3D and SH model update from these captures, the iterative second capture trajectory was able to seek out a new set of unique views to improve off the first. The trajectories and view selection from this can be seen in Figure 5.

Photometric Reconstruction Changes

The held-out PSNR maps show different improvement patterns for the two image sets. Images from the heuristic policy produce a small, broadly distributed PSNR increase across much of the viewing sphere. Images from Splat-RL produce larger localized improvements in regions that initially exhibited high reconstruction error, including the region associated with the omitted solar



(a) Coverage point locations.



(b) Polar trajectory above mean PSNR (left) and 3D trajectory snapshot (right).

Figure 3: Coverage and trajectory associated with the **heuristic** coverage policy.

panel. The Splat-RL refinement also creates some localized regions in which the PSNR error increases substantially, which indicates that a small update set can improve the target defect while worsening error elsewhere. Figure 7 shows the initial heuristic-updated model, Splat-RL single trajectory updated model, and the Splat-RL double trajectory model in one common scale so that these effects remain directly comparable.

Geometric Reconstruction

The most noticeable qualitative result focuses on the intentionally omitted solar panel. Refining the coarse model with the heuristic-policy images does not reconstruct the missing panel since the policy never collects the critical near-eclipse views, and the views it does capture contain intense shadows. Refining the same coarse model with the Splat-RL images adds critical information about the missing-panel region, completing the missing geometry. A second iteration from RL-Splat further increases performance and reduces overfitting from the first 31 images alone. Figure 8 shows the initial, heuristic-updated, and Splat-RL-updated depth or geometry products side by side.

The preliminary geometric metrics favor Splat-RL across all reported measures. The Splat-RL reconstruction achieves a lower Chamfer distance, and a lower Hausdorff distance than the model refined with heuristic-policy images. These improvements support the interpretation that the Splat-

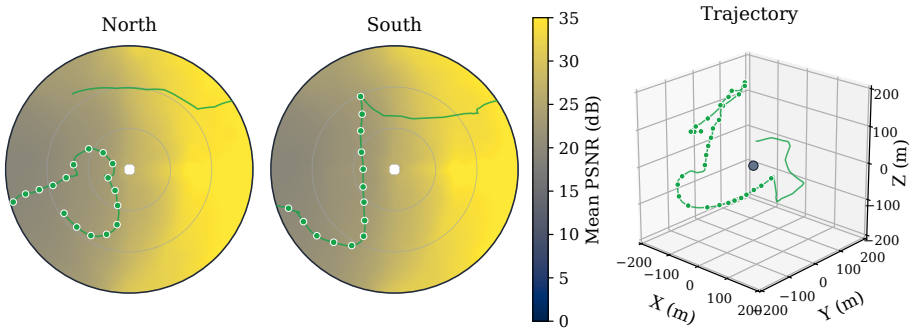
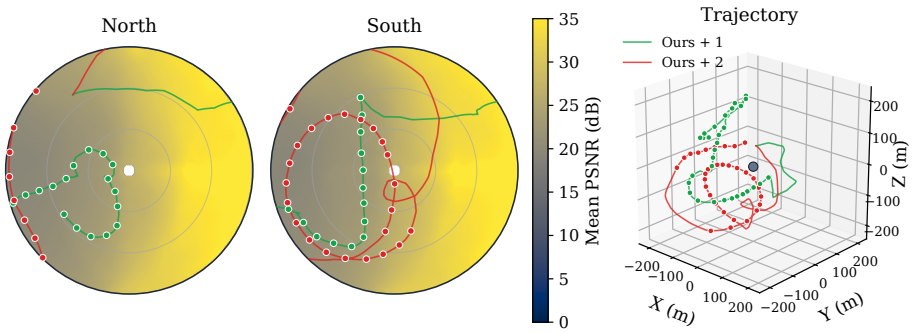


Figure 4: Coverage and trajectory associated with the **Splat-RL** policy for **one trajectory**.



(a) Coverage and image-capture locations.

Figure 5: Coverage and trajectory associated with the **Splat-RL** policy for **two trajectories**.

RL image set contributes more useful geometric information under the shared 31-image budget. Table 1 shows the final values for 3D metric results.

DISCUSSION

Recovering Complex Geometry

The missing-panel result illustrates a structural limitation of coverage-based inspection objectives. The baseline only rewards viewpoints that correspond to unobserved spherical waypoints and satisfy its fixed illumination rule. For a nadir-pointing spacecraft in LEO, a large set of geometrically useful views can occur near eclipse without placing the relevant surface normal within 60° of the Sun. Additionally, since the coverage heuristic neglects complex shadowing effects, the views it does seek are not guaranteed to have sufficient illumination for accurate reconstruction. Splat-RL makes no assumption about surface lighting conditions, since it seeks out views based on reconstruction error. The reward map reflects the combined effect of the true target geometry, the available coarse-model training images, the renderer, and the current reconstruction. This flexibility lets the policy revisit the missing-panel region even though a coverage rule would label the associated waypoint invalid. This comparison demonstrates that our chosen objective design seeks out informative and well-illuminated views without explicit constraints on RSO illumination.

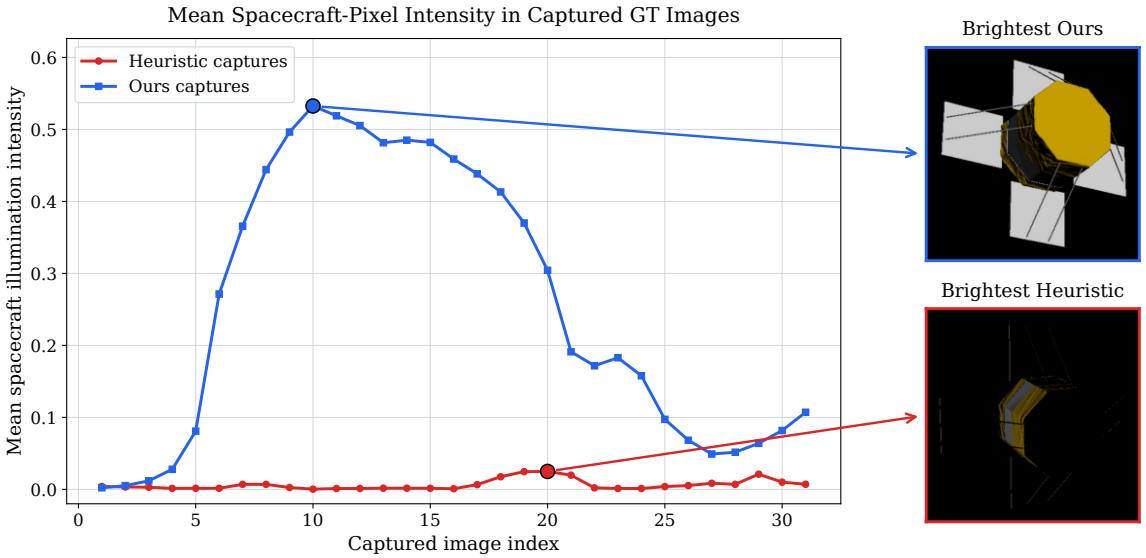


Figure 6: Curves showing the overall illumination for each image taken by a Splat-RL and heuristic policy. The best case illumination for each is displayed on the right.

Table 1: Geometric reconstruction metrics after matched refinement with policy-selected images. Chamfer and Hausdorff distances are normalized by the characteristic solar-panel span of the spacecraft model and are therefore dimensionless. Lower values indicate better reconstruction.

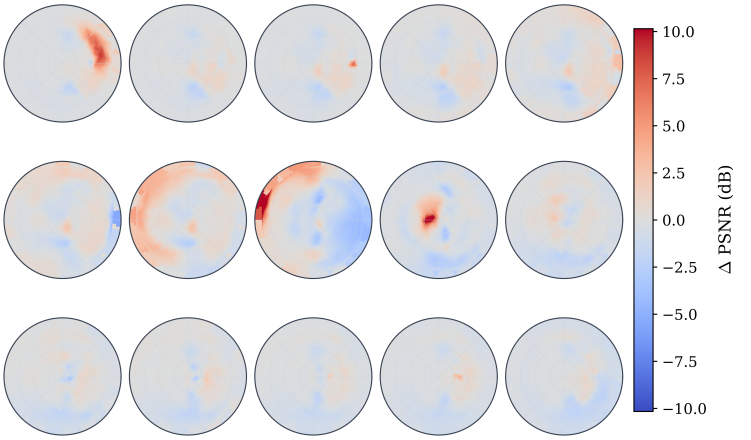
Policy	Chamfer ↓	Hausdorff-95 ↓	Hausdorff Abs. ↓
Heuristic coverage	0.0704	0.1696	0.3008
Splat-RL, one trajectory	0.0394	0.0789	0.2073
Splat-RL, two trajectories	**0.0217**	**0.0302**	**0.1075**

Interpretation of Reconstruction Error

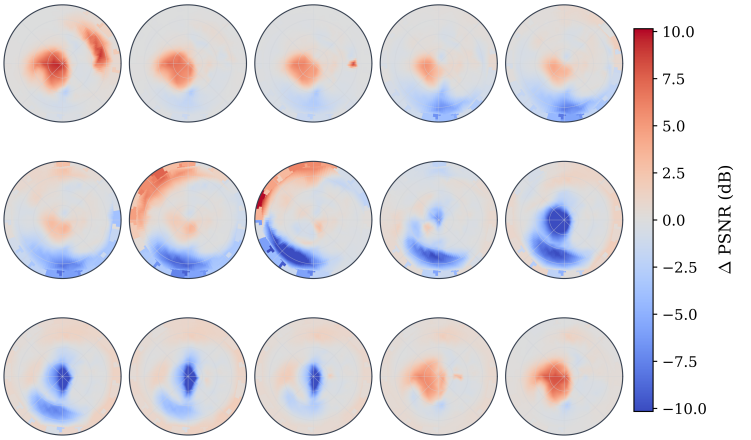
PSNR provides an efficient signal for these experiments because the simulation can render matched GS and ground-truth images across many view and Sun conditions. This signal directly identifies directions in which the coarse model fails to reproduce the RSO's appearance. Our experiments show that this error can guide the policy toward views that improve poorly learned 3D geometry from strictly 2D inputs. Our photometric results also show that localized improvement can accompany degradation elsewhere. Splat-RL produces stronger gains in targeted regions, but includes large losses in others. This behavior may result from the small update set, limited optimization, densification changes, or conflicting appearance supervision across Sun conditions. This may also reflect our GS models current limitations in relighting scenes, which currently requires dense image sampling across many lighting conditions.

Scope and Limitations

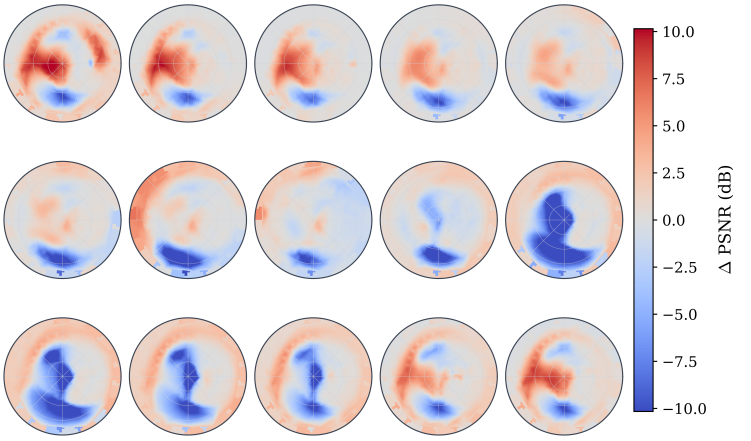
The experiment deliberately removes fuel optimality from the comparison. Both policies retain a 0.1 m/s per-decision impulse limit, collision boundary, keep-in sphere, and episode-duration limit,



(a) Change in mean PSNR after refinement with images selected by the **heuristic coverage policy**.



(b) Change in mean PSNR after refinement with images selected by **single trajectory Splat-RL**.



(c) Change in mean PSNR after refinement with images selected by **double trajectory Splat-RL**.

Figure 7: Held-out PSNR changes, excluding eclipsed snapshots, relative to the initial crude GS model. Each polar plot is a mapping of the view dependent change in reward around the RSO for that sun angle, with the north pole at the center and the south pole at the edge.

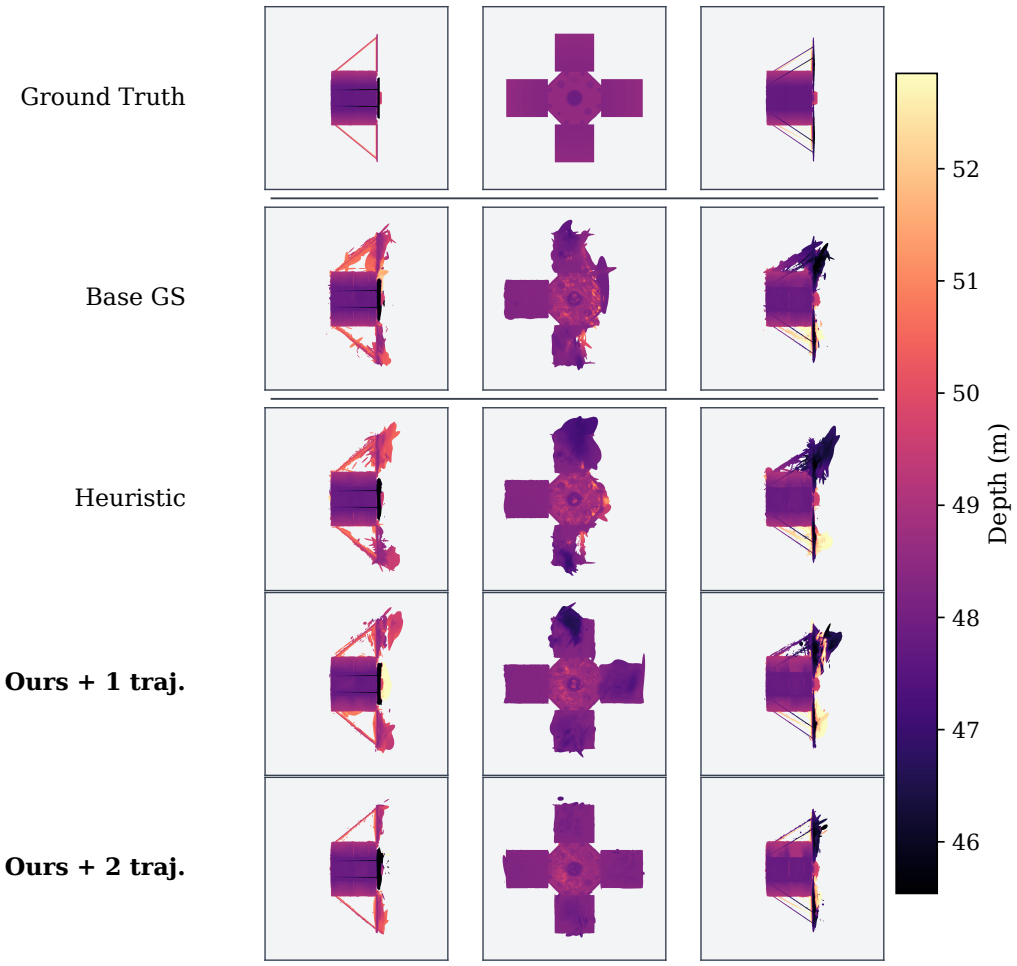


Figure 8: Depth-map comparison of the initial coarse GS model and the models refined with images selected by the heuristic coverage policy and Splat-RL. The heuristic-selected images fail to recover the missing solar panel, whereas the Splat-RL-selected images produce additional reconstructed depth in the corresponding region.

but neither policy receives a cumulative ΔV constraint or penalty. This differs from the full Stephenson et al. formulation, which studies fuel-efficient inspection and safety guarantees in addition to coverage. The present design prevents fuel weighting from obscuring the failure mode caused by the inspection objective. The resulting trajectories should therefore be interpreted as matched-budget sensing trajectories rather than operationally optimal flight plans.

The simulation also simplifies the orbital and sensing problem. It assumes a circular, zero-inclination orbit, a nadir-locked target, exact relative state knowledge, impulsive translational control, and a camera that always points toward the target center. The study omits independent attitude motion, camera pointing constraints, pose-estimation error, orbital perturbations, actuator limits beyond the per-step impulse bound, and explicit safety shielding. The initial defect arises from a designed omission in a known ACRIMSAT dataset, and both policies train and evaluate on that controlled target. These assumptions isolate reconstruction-aware view selection but more work is

needed to approach a fully autonomous solution on unknown RSO's.

The current experiment also performs GS updates after each policy rollout rather than inside the rollout. This structure tests whether the initial reconstruction state can produce a better first acquisition batch, but it does not yet demonstrate repeated onboard adaptation. A fully closed-loop Splat-RL system would refine the GS model, recompute or incrementally update the error representation, and continue planning as the data product changes. Repeated optimization remains computationally expensive because each reconstruction update can require rendering, GS training, and error-model fitting. Future work must determine update schedules and compact model architectures that satisfy onboard compute and memory limits.

Future Work

To eventually reach fully autonomous inspection of an unknown RSO, our work needs to be expanded to account for numerous training and environmental problems. Splat-RL must be expanded to handle a wider domain of orbital conditions outside of the purely circular LEO environment displayed here. It must also be exposed to tighter mission constraints to ensure that we achieve fuel efficient and safe trajectories when inspecting RSO's. Another significant issue is the ability to update GS models within a rollout rather than between rollouts, largely due to the extreme computational demands of training multiple RL and GS models simultaneously. Future work should focus on methods to lighten this process, such that the RL agent can take advantage of both environment dynamics and GS update behavior.

Future reconstruction objectives should also move beyond PSNR alone. A learned predictor could estimate expected improvement in held-out PSNR, depth error, Chamfer distance, or another task-specific metric after a capture. Policies that receive both photometric and geometric uncertainty could then distinguish appearance errors from missing structure. More expressive directional representations could also preserve sharp error features while retaining a manageable observation size. These extensions would move Splat-RL from a 2D reconstruction-error based reward toward a direct 3D active-perception policy for autonomous onboard model building.

CONCLUSION

This work introduces Splat-RL, a reconstruction-aware framework for autonomous spacecraft inspection. Our framework converts the directional PSNR error of a coarse GS model into a compact Sun-conditioned spherical-harmonic representation, which is supplied to a PPO inspection policy. A matched comparison against a coverage heuristic policy gives both policies 31 imaging opportunities and uses their selected images to refine identical copies of the same incomplete GS model. While the coverage policy completes its admissible spherical objective, it fails to reconstruct an intentionally omitted solar panel. The primary factors for this failure included the fixed illumination rule rejecting critical near-eclipse viewpoints, and the spherical geometry approximation causing resulting observations to neglect critical shadowing effects. Splat-RL visits these high-error regions, recovers missing structure, and produces better preliminary three-dimensional metrics across all reported measures. These results support reconstruction-derived objectives as a promising alternative to coverage proxies, while future work must establish statistical reliability, fuel-aware performance, closed-loop reconstruction updates, and generalization to more realistic orbital and sensing conditions.

DISCLOSURE

Some editorial restructuring and language refinement in this manuscript were performed with ChatGPT.

REFERENCES

[1] A. Richards, T. Schouwenaars, J. P. How, and E. Feron, "Spacecraft trajectory planning with avoidance constraints using mixed-integer linear programming," *Journal of Guidance, Control, and Dynamics*, vol. 25, no. 4, pp. 755–764, 2002.

[2] J. A. Starek, E. Schmerling, G. D. Maher, B. W. Barbee, and M. Pavone, "Fast, safe, propellant-efficient spacecraft motion planning under clohessy–wiltshire–hill dynamics," *Journal of Guidance, Control, and Dynamics*, vol. 40, no. 2, pp. 418–438, 2017.

[3] S. Faghihi, S. Tavana, and A. H. J. de Ruiter, "Kinodynamic on-orbit inspection path planning for full-coverage inspection in close proximity of space structures," *Acta Astronautica*, vol. 198, pp. 354–365, 2022.

[4] ——, "Multiple spacecraft coordination and motion planning for full-coverage inspection of large complex space structures," *Acta Astronautica*, vol. 202, pp. 119–129, 2023.

[5] F. Capolupo and P. Labourdette, "Receding-horizon trajectory planning algorithm for passively safe on-orbit inspection missions," *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 5, pp. 1023–1032, 2019.

[6] M. Maestrini and P. Di Lizia, "Guidance strategy for autonomous inspection of unknown non-cooperative resident space objects," *Journal of Guidance, Control, and Dynamics*, vol. 45, no. 6, pp. 1126–1136, 2022.

[7] Y. K. K. Nakka, W. Hönig, C. Choi, A. Harvard, A. Rahmani, and S.-J. Chung, "Information-based guidance and control architecture for multi-spacecraft on-orbit inspection," *Journal of Guidance, Control, and Dynamics*, vol. 45, no. 7, pp. 1184–1201, 2022.

[8] M. Hibbard, M. Cubuktepe, M. Shubert, K. Lang, U. Topcu, and S. Phillips, "Trajectory synthesis for the coordinated inspection of a spacecraft with safety guarantees," *Journal of Guidance, Control, and Dynamics*, vol. 46, no. 12, pp. 2245–2264, 2023.

[9] K. Hovell and S. Ulrich, "Deep reinforcement learning for spacecraft proximity operations guidance," *Journal of Spacecraft and Rockets*, vol. 58, no. 2, pp. 254–264, 2021.

[10] L. Federici, B. Benedikter, and A. Zavoli, "Deep learning techniques for autonomous spacecraft guidance during proximity operations," *Journal of Spacecraft and Rockets*, vol. 58, no. 6, pp. 1774–1785, 2021.

[11] D. E. J. van Wijk, K. Dunlap, M. Majji, and K. L. Hobbs, "Deep reinforcement learning for autonomous spacecraft inspection using illumination," in *AAS/AIAA Astrodynamics Specialist Conference*, Big Sky, Montana, 2023, paper AAS 23-342.

[12] M. Stephenson, D. Huterer Prats, and H. Schaub, "Autonomous satellite inspection in low earth orbit with optimization-based safety guarantees," in *International Workshop on Planning & Scheduling for Space*, Toulouse, France, April 28–30 2025, available at https://hanspeterschaub.info/Papers/Stephenson2025a.pdf.

[13] K. Dunlap, K. Bennett, D. E. J. van Wijk, N. Hamilton, and K. L. Hobbs, "Run time assured reinforcement learning for six-degree-of-freedom spacecraft inspection," in *AIAA AVIATION FORUM AND ASCEND 2024*, 2024, aIAA Paper 2024-4935.

[14] A. Brandonisio, L. Capra, and M. Lavagna, "Deep reinforcement learning spacecraft guidance with state uncertainty for autonomous shape reconstruction of uncooperative target," *Advances in Space Research*, vol. 73, no. 11, pp. 5741–5755, 2024.

[15] J. Aurand, S. Cutlip, H. Lei, K. Lang, and S. Phillips, "Deep q-learning for decentralized multi-agent inspection of a tumbling target," *Journal of Spacecraft and Rockets*, vol. 61, no. 2, pp. 341–354, 2024.

[16] S. Lee, L. Chen, J. Wang, A. Liniger, S. Kumar, and F. Yu, "Uncertainty guided policy for active robotic 3d reconstruction using neural radiance fields," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12 070–12 077, 2022.

[17] X. Chen, Q. Li, T. Wang, T. Xue, and J. Pang, "GenNBV: Generalizable next-best-view policy for active 3d reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 436–16 445.

[18] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104–4113.

[19] R. W. Gaskell, O. S. Barnouin, M. G. Daly, E. E. Palmer, J. R. Weirich, C. M. Ernst, R. T. Daly, and D. S. Lauretta, "Stereophotoclinometry on the osiris-rex mission: mathematics and methods," *The Planetary Science Journal*, vol. 4, no. 4, p. 63, 2023.

[20] T. Driver, A. Vaughan, Y. Cheng, A. Ansar, J. Christian, and P. Tsiotras, "Stereophotoclinometry revisited," *arXiv preprint arXiv:2504.08252*, 2025.

[21] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision*, 2020.

[22] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM Transactions on Graphics*, vol. 42, no. 4, pp. 1–14, 2023.

[23] V. M. Nguyen, E. Sandidge, T. Mahendrakar, and R. T. White, "Characterizing satellite geometry via accelerated 3d gaussian splatting," *Aerospace*, vol. 11, no. 3, p. 183, 2024.

[24] T. H. Park and S. D'Amico, "Improved 3d gaussian splatting of unknown spacecraft structure using space environment illumination knowledge," 2025. [Online]. Available: https://arxiv.org/abs/2512.23998

[25] A. Ghosh, S. Achutha, W. Heidrich, and M. O'Toole, "BRDF acquisition with basis illumination," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[26] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*. ACM Press/Addison-Wesley Publishing Co., 2000, pp. 145–156.

[27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[28] M. Tancik, E. Weber, E. Ng, R. Li, B. Yi, T. Wang, A. Kristoffersen, J. Austin, K. Salahi, A. Ahuja, D. Mcallister, J. Kerr, and A. Kanazawa, "Nerfstudio: A modular framework for neural radiance field development," in *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings*, ser. SIGGRAPH '23. ACM, 2023, p. 1–12. [Online]. Available: http://dx.doi.org/10.1145/3588432.3591516