Assessing the Predictive Value of Physics-Grounded Synthetic Data for Computer Vision in Space Environments

Anonymous CVPR submission

Paper ID *****

Abstract

Synthetic imagery is widely used to train space-domain vision systems for space applications because real orbital imagery is scarce and expensive to collect, yet it remains unclear which aspects of physical realism in synthetic data generation influence sim-to-real generalization. In this work, we investigate the predictive value of physics-grounded synthetic data using a Digital Twin Environment (DTE) that generates spacecraft imagery under controlled orbital geometry while varying illumination and surface reflectance modeling. Rather than optimizing downstream models, we fix the learning pipeline and use detection performance as a probe of data quality. Models trained under different synthetic regimes are evaluated on a web-scraped spacecraft imagery dataset and on a real laboratory dataset of an unseen spacecraft mockup under space-like lighting. Synthetic-only training fails to transfer to real imagery, achieving near-zero detection and segmentation accuracy on the laboratory dataset. However, when synthetic imagery is used to augment web-collected training imagery, performance improves substantially, with the best model reaching 0.724 mAP@0.5 on the laboratory dataset, improving upon previously reported performance on this dataset (0.394 mAP@0.5). To interpret these differences, we analyze internal feature representations using PEEK-based information statistics and find that synthetic-only training concentrates representational variance near the backbone terminus, whereas models trained with real-data involvement maintain more distributed deep representations. Rotation estimation experiments using Structure-from-Motion reconstruction further show that physics-grounded rendering is necessary when synthetic imagery is intended to emulate real-world observations, as heuristic appearance models fail to preserve feature correspondences required by motion-estimation pipelines. These results provide insight into how synthetic data influences representation learning and offer guidance for designing synthetic data pipelines for space-domain vision systems.

1. Introduction

Vision systems for space operations require high reliability, yet real training imagery of spacecraft under operational scenarios is scarce, difficult to collect, and expensive to annotate. Missions involving inspection, docking, and on-orbit servicing require perception capabilities that function robustly despite extreme illumination conditions, large phase-angle variations, and complex spacecraft geometries with highly reflective surfaces, where perception failures can jeopardize mission safety. The combination of limited real data and stringent reliability requirements has made synthetic imagery an essential tool for developing and evaluating space-domain vision systems.

However, the increasing reliance on synthetic imagery raises an important question: *which aspects of physical realism in synthetic data generation influence sim-to-real transfer?* Many pipelines rely on heuristic lighting, simplified motion models, or visually plausible textures that may not preserve the visual cues required for models trained on synthetic data to generalize to real observations.

In this work, we investigate the predictive value of physics-grounded synthetic data for space-domain vision. We use a Digital Twin Environment (DTE) [5, 6] that integrates orbital geometry, ephemeris-driven illumination, and bidirectional reflectance distribution function (BRDF) appearance modeling to generate spacecraft imagery under controlled conditions. Rather than optimizing downstream models, our goal is to evaluate how variations in the physical realism of synthetic imagery influence sim-to-real generalization.

To isolate the effects of synthetic data generation, we adopt a fixed learning pipeline and treat downstream vision models as measurement probes. Object detection and instance segmentation models are trained under multiple data regimes, including synthetic-only training, training on web-scraped spacecraft imagery, and the same augmented with synthetic imagery. Across these experiments, synthetic image regimes we generate differ only in how physical factors such as illumination and surface reflectance are modeled.

Models are evaluated on a held-out portion of the web-scraped spacecraft imagery dataset and on real laboratory mockup, enabling controlled measurement of cross-dataset generalization and sim-to-real transfer.

Beyond task-level performance, we analyze internal feature representations using the PEEK framework [10], which summarizes how activation entropy and its spatial variance is distributed across network layers when models process real imagery and provides a compact measure of how representational variance is distributed across network depth.

Using the DTE, we generate a large synthetic spacecraft dataset with automatically generated detection and segmentation annotations derived from CAD geometry. These datasets allow illumination and reflectance modeling to be varied independently of orbital geometry and motion.

This work makes the following contributions:

- We extend a physics-grounded DTE capable of generating spacecraft imagery with controllable orbital geometry, ephemeris-driven illumination, and BRDF-based surface appearance.
- We introduce a controlled experimental methodology that evaluates the predictive value of synthetic data by fixing the learning pipeline and varying only the physical realism of rendered imagery.
- We perform a sim-to-real evaluation of spacecraft component detectors trained under multiple data regimes, including synthetic-only training and real datasets augmented with synthetic imagery.
- We release mirrored synthetic spacecraft datasets, each containing 139,212 images across three spacecraft models with automatically generated detection and segmentation annotations, depth maps, and camera poses, rendered under identical orbital configurations but using heuristic and physics-grounded appearance models.

2. Related Work

Synthetic imagery has become a central tool for training vision systems for spacecraft perception due to the scarcity of labeled orbital imagery. Early work demonstrated that convolutional neural networks trained on rendered imagery can estimate the pose of a known noncooperative spacecraft from monocular images [24]. The associated SPEED dataset and its successor SPEED+ introduced large-scale synthetic imagery together with hardware-in-the-loop captures to study the domain gap between simulated and real observations [18, 21]. Subsequent research explored strategies for improving robustness across domains, including multi-task learning and online domain refinement techniques designed to adapt models trained on synthetic data to real imagery [19, 20]. More recent work has emphasized physically based rendering pipelines that incorporate realistic materials, lighting, and environmental effects in order to improve visual fidelity and reduce the sim-to-real gap [5, 6, 26].

Spacecraft perception research has also focused on detecting and segmenting spacecraft components to support proximity operations and on-orbit servicing tasks. Public datasets containing spacecraft images with bounding boxes and part-level annotations have enabled the evaluation of object detection and instance segmentation models for space applications [3]. Region-based convolutional networks and related architectures have been used to identify spacecraft components under challenging viewing geometries and illumination conditions [2, 7]. Because large real-world datasets are difficult to obtain, several studies have relied on automatically generated synthetic imagery to train perception models for noncooperative resident space objects [4, 27]. Lightweight detection architectures suitable for onboard deployment have further demonstrated that modern object detectors can identify spacecraft components such as solar panels, antennas, and thrusters even under previously unseen orientations [8, 9]. These works motivate the use of detection and segmentation tasks as practical probes for evaluating the quality of synthetic training data.

Beyond task-level performance metrics, recent work has examined the internal representations learned by spacecraft perception models. Entropy-based analyses of intermediate activations have been used to provide probabilistic explanations for satellite component detectors [10, 14] and pruning them for efficiency [12]. Building on this line of research, information-theoretic analysis of neural representations introduced PEEK variance, a measure of spatial information diversity computed from feature activations that links representational structure to optimization behavior [10]. Aggregating these statistics across layers yields the Relative Variance Contribution (RVC), which quantifies the spatial informational diversity of each layer. These representation-level metrics provide complementary insight into how neural networks process visual information and offer additional tools for analyzing how different training data regimes influence learned representations.

Physically grounded simulation environments have recently been proposed to generate synthetic spacecraft imagery under realistic orbital observation conditions. We adopt and upgrade the DTE introduced in [5, 6], which models orbital geometry and illumination to support spacecraft observation and reconstruction tasks. However, most prior work focuses either on generating synthetic datasets or on improving learning algorithms to mitigate domain shift. Comparatively little work has examined how specific rendering assumptions within synthetic data pipelines influence downstream perception performance. To address this gap, this paper extends this physics-grounded DTE to systematically compare heuristic and physically grounded rendering regimes under identical orbital conditions. By isolating the effects of lighting models and surface reflectance

in synthetic imagery, we evaluate how simulation fidelity affects generalization across synthetic domains and transfer to laboratory observations.

3. Physics-Grounded Digital Twin Environment

The DTE provides a controllable simulation framework for generating spacecraft imagery under physically consistent orbital conditions. A key design goal of the DTE is to separate scene geometry from rendering assumptions so that physical factors affecting spacecraft appearance can be varied while all other scene parameters remain fixed. This enables the generation of mirrored synthetic datasets in which spacecraft geometry, trajectories, and camera configurations are identical while illumination and surface appearance models differ. Figure 1 illustrates the synthetic data generation pipeline. Orbital parameters, spacecraft models, and rendering configurations are specified through a configuration interface that defines the relative geometry between an observing spacecraft and a resident space object (RSO), illumination conditions determined by orbital ephemerides, and the physical surface properties used for rendering. The environment then produces synchronized image sequences together with object-level bounding box and mask annotations.

Geometry and Orbital Context. RSO trajectories are defined using classical orbital elements in an inertial reference frame. Relative motion between the observing spacecraft and the RSO is represented in the local vertical local horizontal (LVLH) frame using linearized orbital dynamics (Clohessy–Wiltshire equations), enabling simulation of inspection trajectories such as V-bar approaches and natural motion circumnavigation (NMC). Camera poses are derived directly from the simulated spacecraft state to ensure geometric consistency between orbital configuration and rendered imagery.

Appearance and Illumination Modeling. The visual appearance of spacecraft surfaces is modeled using BRDFs parameterized by the Hapke reflectance model, which captures the directional dependence of surface reflectance under varying lighting and viewing angles. Illumination conditions are derived from orbital ephemerides that determine the position of the Sun relative to the spacecraft at a given time. This combination allows the simulator to reproduce phase-angle effects, moving shadows, and eclipse intervals that naturally occur in orbit. For the experiments in this work, mirrored synthetic datasets are rendered under identical orbital configurations using two appearance regimes: a heuristic rendering model and a physics-grounded configuration incorporating ephemeris-driven illumination and Hapke-based surface reflectance.

Motion Modeling. The DTE supports rigid-body rotation of the RSO, allowing simulation of tumbling or rotating spacecraft in which both viewpoint and spacecraft orientation evolve over time.

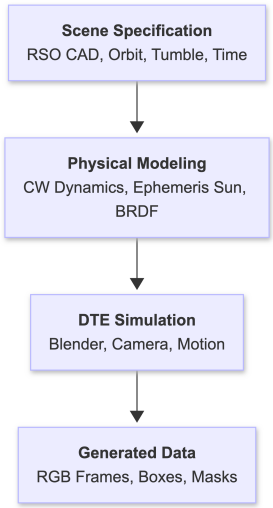

Figure 1. Overview of the physics-grounded Digital Twin Environment (DTE) and synthetic data generation pipeline. The DTE integrates orbital context, ephemeris-driven illumination, BRDF-based appearance modeling, and spacecraft motion to produce synthetic imagery with corresponding object-level annotations.

Synthetic Annotation Generation. Spacecraft CAD models [15–17] are decomposed into semantically meaningful components such as antennas, solar arrays, and bus structures. These components are registered within the rendering environment so that object-level annotations can be generated automatically during rendering. The resulting images include aligned detection bounding boxes and pixel-wise segmentation masks without manual labeling.

Reproducibility and Control. All simulation parameters are specified through deterministic configuration files, allowing rendering conditions to be reproduced exactly across experiments. Orbital geometry, trajectories, spacecraft models, illumination timing, and rendering assumptions can be varied independently, enabling controlled generation of synthetic datasets for systematic evaluation.

Because the mirrored datasets share identical spacecraft geometry, trajectories, and camera configurations, differences in the rendered imagery arise solely from illumination and surface appearance modeling. Figure 2 shows example mirrored renders generated from identical spacecraft states and camera poses under heuristic and physics-grounded appearance models, outlining key visual features reproduced only in the latter.

4. Validation Methodology

Our objective is to evaluate the predictive value of synthetic data generated by the DTE rather than optimize downstream

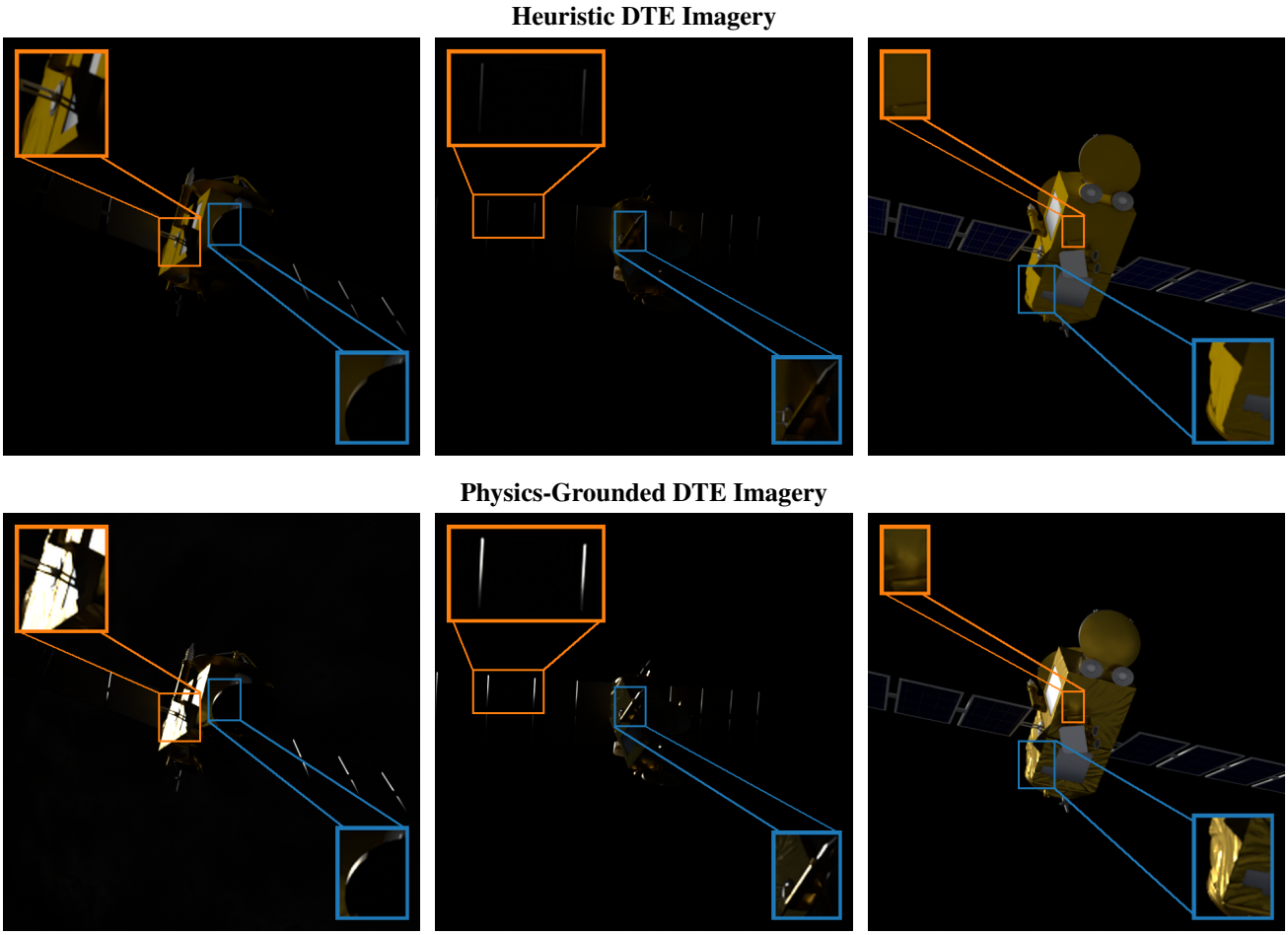

Figure 2. Comparison of heuristic synthetic imagery (top row) and physics-grounded rendering (bottom row) under identical spacecraft pose and camera configuration. Columns highlight representative visual effects: (left) harsh illumination conditions, (middle) reflectance behavior in shadowed regions, and (right) surface texture and reflection differences.

vision models. We adopt a fixed learning pipeline and vary only the training data composition and rendering assumptions. Downstream models are treated as measurement probes whose performance reflects how well synthetic imagery captures visual cues that transfer to real observations.

Downstream Models as Measurement Probes. Object detection and instance segmentation serve as probe tasks because they require models to localize and distinguish spacecraft components under varying illumination and viewing geometry. Differences in performance across training regimes therefore indicate how well synthetic imagery captures physically meaningful visual structure relevant for real observations.

Geometric Probe via Rotation Estimation. Rotation angle and rate estimation provide a geometric probe because motion-estimation pipelines rely on stable feature correspondences across image sequences. Differences in reconstruction performance across rendering regimes therefore indicate whether synthetic imagery preserves the visual consistency required by geometric vision algorithms. This experiment evaluates whether DTEs intended to emulate real observation sequences require physics-grounded rendering in order to produce imagery suitable for testing such perception systems prior to deployment.

Fixed Learning Pipeline. All experiments use identical preprocessing, training, and inference configurations so that performance differences can be attributed to training data rather than model design. We train YOLO26 [25] detectors and segmenters. Object detectors are trained from scratch with the SGD optimizer, batch size 32, initial learning rate 0.01, and momentum 0.75. Instance segmentation models are trained from scratch with the Adam optimizer, batch size 16, initial learning rate 0.001, and momentum 0.5.

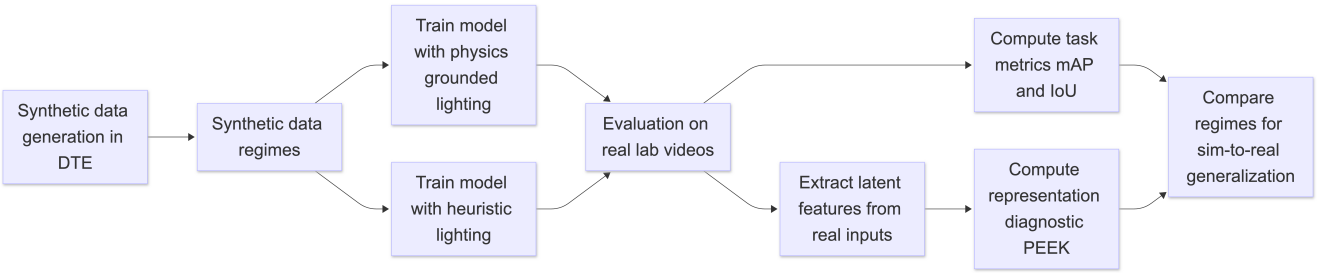

Figure 3. Pipeline for evaluating the predictive value of synthetic data using fixed-model probes and mirrored rendering regimes.

Training and Evaluation Regimes. We evaluate three training regimes: models trained only on synthetic imagery, models trained on web-collected imagery, and models trained on web-collected imagery augmented with synthetic data.

Our evaluation uses four datasets spanning different rendering assumptions and sensing conditions: two mirrored synthetic datasets generated within the DTE, an external spacecraft imagery dataset, and a real laboratory dataset.

Heuristic Synthetic Dataset. Orbital geometry and inspection trajectories are simulated using classical orbital elements and LVLH relative motion under Clohessy–Wiltshire dynamics.

Physics-Grounded Synthetic Dataset. This dataset uses the same orbital geometry and trajectories as the heuristic dataset but incorporates ephemeris-driven solar positioning and BRDF-based surface reflectance (informed by the Hapke reflectance model). The two datasets therefore share identical geometry and camera motion while differing only in illumination and reflectance modeling.

Synthetic Dataset Specifications. Synthetic imagery is generated for multiple spacecraft models including Jason-1 [16], IceSAT [15], and Clementine [17], although Jason-1 provides the primary synthetic training data used in our experiments due to its more-detailed geometry and component structure. Inspection trajectories at LEO altitudes (719 km) include NMC, corkscrew, coplanar, and V-bar maneuvers with 10–50 m offsets and $1 - 10°$/min rotaton rate. Two mirrored datasets are generated, each containing 139,212 images, with identical orbital configurations but differing in appearance modeling (heuristic vs. physics-grounded).

Web Satellite Dataset (WSD). To evaluate generalization across independent data sources, we include the Web Satellite Dataset (WSD) [8], which contains diverse spacecraft imagery collected from the internet, including diagrams, ground imagery of spacecraft parts, renderings, and real in-space images. Models are trained using the WSD training split and evaluated on the corresponding test set, with synthetic imagery optionally used as augmentation during training.

Laboratory Dataset. To assess sim-to-real transfer, we evaluate models on a laboratory dataset consisting of video imagery of a spacecraft mockup captured under controlled motion and lighting conditions [1]. This dataset is used exclusively for evaluation and is never used during training. It contains 1,800 labeled images corresponding to a full 360° yaw rotation of the spacecraft.

Training Protocol. Models are trained using the synthetic-only, WSD-only, and WSD+synthetic regimes described above and evaluated on both WSD and laboratory imagery.

Evaluation Metrics. We report standard object detection and segmentation metrics including mAP@0.5 and and mAP@0.5:0.95. Performance on the laboratory dataset serves as the primary indicator of sim-to-real transfer, while results on the WSD dataset provide an additional measure of cross-dataset generalization.

Representation-Level Analysis. In addition to task-level performance, we analyze internal model representations using the PEEK framework [10] to examine how training under different data regimes influences the spatial organization of feature activations when models are applied to real imagery.

5. Generalizing across the Domain Gap

We evaluate how training data composition affects transfer from synthetic imagery to real observations. Detection models are trained using the protocols described in Section 4 and evaluated on two domains that were never seen during training: a held-out subset of the WSD and real laboratory imagery of a spacecraft mockup.

Table 1. Detection and segmentation performance across evaluation domains. For training sets, (H) indicates heuristic synthetics and (P) indicates physics-grounded synthetics. All test sets are unseen during training. Box metrics correspond to detection (mAP@0.5 and mAP@0.5:0.95), while Mask metrics correspond to segmentation.

Training Data	**WSD**				**Lab**			
	Box		Mask		Box		Mask	
	mAP50	mAP50–95	mAP50	mAP50–95	mAP50	mAP50–95	mAP50	mAP50–95
WSD	0.588	0.277	0.250	0.145	0.629	0.194	0.062	0.024
Jason-1 (H)	0.068	0.029	0.073	0.035	0.000	0.000	0.000	0.000
Jason-1 (P)	0.056	0.020	0.101	0.044	0.000	0.000	0.000	0.000
WSD + Jason-1 (H)	**0.647**	**0.347**	**0.341**	**0.200**	**0.724**	**0.205**	**0.109**	**0.081**
WSD + Jason-1 (P)	0.582	0.281	0.314	0.183	0.362	0.104	0.090	0.072

Synthetic-only training does not transfer to real imagery. The heuristic-only and physics-only detectors achieve only 0.068 and 0.056 mAP@0.5 on the WSD test set, respectively, with mAP@0.5:0.95 values below 0.03. Instance segmentation models trained on synthetic data alone show similarly limited transfer, achieving mask mAP@0.5 values of 0.073 and 0.101 on WSD. Performance on the laboratory dataset collapses almost entirely for both tasks, with near-zero detection and segmentation accuracy on real laboratory imagery.

Combining WSD with Jason-1 synthetic imagery yields the strongest overall performance. The WSD+Jason1 (heuristic) model achieves the highest performance on the WSD test set with 0.647 mAP@0.5 and 0.347 mAP@0.5:0.95, improving over WSD-only training by 0.059 and 0.070 respectively. A similar pattern holds for instance segmentation, where WSD+Jason1 (heuristic) achieves the highest mask mAP across both evaluation domains. On the laboratory dataset it also achieves the best localization performance among all models, reaching 0.724 mAP@0.5, strongly exceeding previously reported results on this dataset [14] of 0.394, and 0.205 mAP@0.5:0.95. These results indicate that synthetic augmentation improves localization robustness and illumination invariance when paired with real imagery rather than used in isolation.

Instance segmentation performance was uniformly poor across all training regimes and did not meaningfully differentiate between models, so it is not analyzed further.

Synthetic augmentation benefits illumination-sensitive components. The largest gains from synthetic augmentation occur for solar-panel detection. In both standard and low-light laboratory imagery, WSD+Jason1-heuristic substantially improves solar-panel accuracy relative to WSD-only training. By contrast, spacecraft body detection is already strong under WSD-only training, while antennas and thrusters remain challenging across all training regimes due to their small spatial footprint and frequent occlusion. This indicates synthetic augmentation is most beneficial for visually sensitive structures whose appearance varies strongly with illumination and viewing geometry.

Rotation Estimation from Synthetic Imagery. As a complementary experiment related to the domain gap, we evaluated whether heuristic synthetic imagery preserves the visual cues required by geometric motion-estimation pipelines. In each synthetic data regime, image sequences from simulated Jason-1 V-bar inspection trajectories (10 m offset) with tumbling motion (1°/min and 5°/min about the principal axis) were processed using Hierarchical Localization (HLOC) [22] followed by COLMAP [23] with hyperparameters from prior spacecraft reconstruction work [5]. COLMAP reconstructs a sparse point cloud and camera trajectory assuming a static scene; spacecraft rotation is therefore absorbed into the estimated camera motion.

Spacecraft rotation is recovered by comparing the reconstructed camera trajectory with the known inspection trajectory expressed in the LVLH frame. At 5°/min tumbling, both renderings yield viable Structure-from-Motion (SfM) reconstructions, but the physics-grounded imagery reconstructs 2559 3D points compared with 1733 for the heuristic rendering, producing rotation rate and angle errors three orders of magnitude smaller. At 1°/min tumbling, the physics-grounded sequence still produces a stable reconstruction, whereas the heuristic sequence fragments into small components and fails to yield reliable motion estimates (Table 2). This reflects the classical SfM tradeoff between parallax and appearance: when inter-frame motion is small, reconstruction depends on repeatable feature correspondences, which are better preserved by physically grounded illumination and reflectance.

Table 2. Rotation estimation from synthetic inspection imagery using HLOC+COLMAP with mean rotation rate error and mean angle error.

Model	Rate	Rotation Error	Angle Error
Jason-1 (P)	5°/min	7.11e-4	5.32e-5
Jason-1 (H)	5°/min	1.05e-1	3.23e-2
Jason-1 (P)	1°/min	2.26e-2	5.88e-2
Jason-1 (H)	1°/min	Fail	Fail

6. Representation-Level Analysis

To analyze how different training regimes influence internal representations, we use the PEEK framework [10, 13], which measures which measures how activation entropy is spatially distributed within intermediate feature maps. We summarize these maps using the variance of PEEK and the RVC, which quantifies the proportion of spatial representational variance attributable to each network layer. These statistics provide a compact summary of how different training regimes organize internal feature representations when models process the same real imagery.

Task-level metrics provide only a coarse view of sim-to-real generalization and do not reveal how synthetic training regimes shape internal feature representations. PEEK-based statistics therefore provide a complementary diagnostic by summarizing how information is spatially organized within intermediate feature maps. By fixing the evaluation dataset and comparing representation statistics across models trained under different regimes, these metrics allow differences in learned representation structure to be analyzed independently of task accuracy. All analyses are performed on identical held-out real images, ensuring that representational differences arise from training regime rather than input variation.

Qualitative PEEK visualizations provide insights into model focus. Figure 4 shows PEEK maps for the same real laboratory image across models trained under different regimes at two modules near the detection heads. At Module 19, immediately preceding the second detection head, the feature-map scale corresponds to objects occupying roughly one-tenth of the image, closely matching the apparent spacecraft size. Although the synthetic-only models still produce coarse spatial responses near the spacecraft, their high-PEEK responses are diffuse and concentrated near the object and image periphery. In contrast, models trained with real-data involvement produce more structured responses centered on semantically meaningful spacecraft components, particularly the solar panels and antenna.

At the deeper module (Module 21), the Jason-1–only trained models show similar localization behavior, but the WSD-only model shifts its strongest responses toward background structure. In contrast, the hybrid WSD+Jason-1 model remains well localized on the spacecraft, suggesting augmentation with even heuristic synthetics stabilizes deeper representations. Some responses also appear along the laboratory gantry supporting the mock spacecraft, reflecting regular structure present in the lab environment that would not occur in on-orbit imagery. These visualizations suggest that stronger-performing models exhibit more consistent semantic alignment between PEEK responses and spacecraft components, with the hybrid WSD+Jason-1 model showing the most stable localization across modules, while synthetic-only models produce diffuse responses that are weakly aligned with the satellite components.

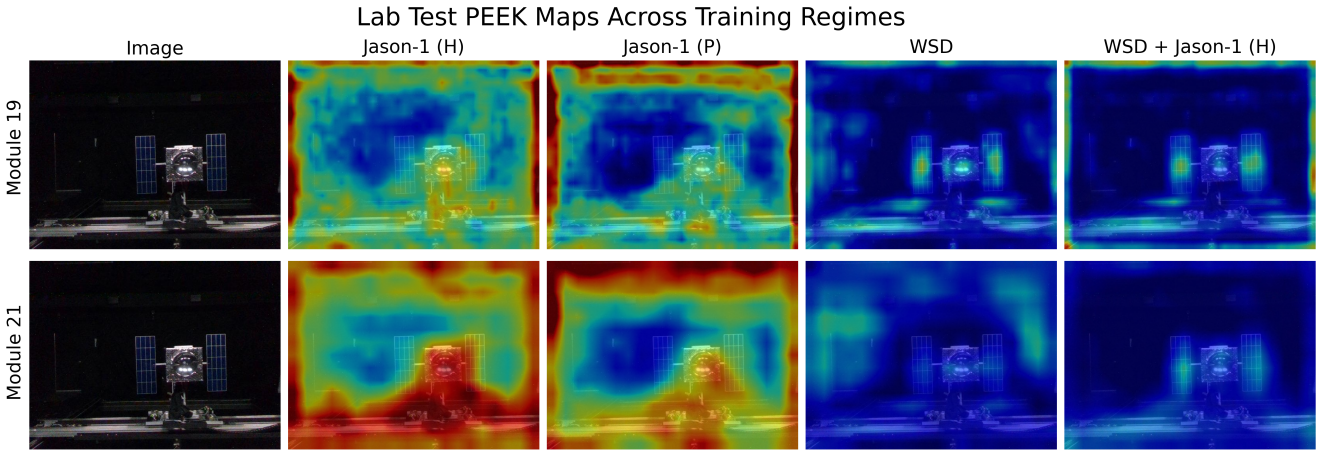

Figure 4. PEEK maps for the same laboratory image across models trained under different regimes. Rows correspond to modules near the detection heads (Modules 19 and 21). Synthetic-only models exhibit diffuse peripheral responses, whereas models trained with real-data involvement focus on spacecraft components. At the deeper module, the WSD-only model shifts attention toward background structure, while the hybrid WSD+Jason-1 model remains well localized.

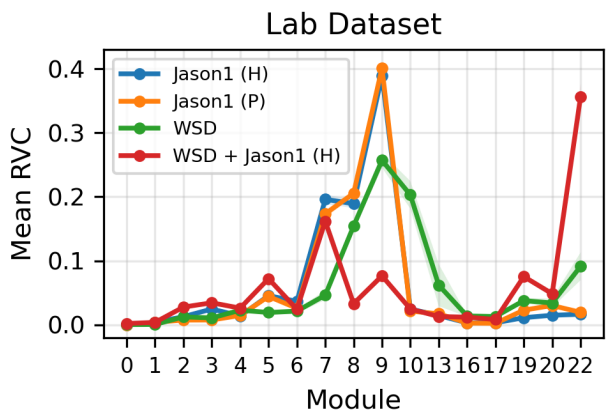

Figure 5. Layer-wise Relative Variance Contribution (RVC) on the real laboratory test sets, averaged across the dataset with 95% confidence interval bands.

Figure 5 shows the layer-wise RVC profiles on the real laboratory dataset for object detectors. The two models trained purely on synthetic data exhibit a pronounced spike at Module 9, the Spatial Pyramid Pooling–Fast (SPPF) block, and concentrate nearly all representational variance in Modules 7-9 near the end of the backbone. In contrast, the far more accurate models trained with real-data involvement show a less dominant SPPF spike and instead have elevated variance in late modules close to the second and third detection heads. This is consistent with object detection failure modes in the literature on RVC [11].

Notably, the two fully synthetic-trained models also perform extremely poorly on the real laboratory images. This suggests the strong concentration of representational variance near the backbone terminus is associated with failed sim-to-real transfer, while models trained with more diverse imagery maintain more distributed deep representations when processing real inputs.

7. Discussion and Conclusions

This study examined which aspects of physical realism in synthetic spacecraft imagery influence sim-to-real transfer across both semantic perception tasks and geometric motion estimation. Our results show that synthetic imagery alone does not bridge the domain gap to real observations, even when generated under physically consistent orbital geometry and lighting. However, synthetic data can significantly improve performance when used as augmentation alongside diverse web-collected data: it improved detection and segmentation performance and increased robustness to variations in viewing geometry and lighting conditions, even when perfect physical realism is not achieved.

Rotation-estimation experiments based on SfM reconstruction further show that physics-grounded rendering is necessary when synthetic imagery is intended to emulate real-world observations, as heuristic appearance models fail to preserve the feature correspondences required by motion-estimation pipelines.

Our representation-level analysis provides additional insight into these results. YOLO26 models trained purely on synthetic imagery concentrate most representational variance (PEEK variance) near the end of the backbone and perform poorly when evaluated on real imagery. In contrast, models trained with WSD or mixed-data regimes maintain more distributed representational variance across deeper layers closer to the detection heads. This suggests successful sim-to-real transfer is associated not only with task-level accuracy but also with how representational variance is distributed across network depth.

These findings indicate that efficiently generated heuristic synthetic imagery can still be useful when used to augment small, web-collected datasets such as WSD, particularly for introducing diverse viewing geometries and lighting conditions. In this setting, geometric diversity contributes more to generalization than physically exact appearance modeling.

This work has several limitations. The evaluation focuses on object detection, instance segmentation, and rotation estimation tasks on synthetic and laboratory imagery and does not consider onboard deployment constraints or runtime performance. Although the experiments use a "small" YOLO detector representative of edge-deployable architectures, the system has not yet been evaluated in a fully embedded flight environment. In addition, the laboratory dataset represents a controlled environment and does not capture the full variability of on-orbit observations.

In conclusion, we extend a physics-grounded DTE and introduce a controlled experimental framework for evaluating the predictive value of synthetic spacecraft imagery through sim-to-real generalization experiments. By generating mirrored synthetic datasets that share identical orbital geometry, camera trajectories, and spacecraft motion while differing only in rendering assumptions, our study isolates how physical realism in appearance modeling affects downstream detection and segmentation performance and internal representation structure. The results show that synthetic augmentation can improve perception performance, while physics-grounded rendering is required when synthetic imagery is intended to emulate real observations used by geometric vision pipelines. These findings provide practical guidance for the design and validation of synthetic data pipelines for space-domain computer vision systems.

References

[1] M. Attzs, Trupti Mahendrakar, Mackenzie Meni, Ryan White, and Isaac Silver. Comparison of Tracking-By-

Detection Algorithms for Real-Time Satellite Component Tracking. *Small Satellite Conference*, 2023. 5

[2] Yulang Chen, Jingmin Gao, and Kebei Zhang. R-CNN-Based Satellite Components Detection in Optical Images. *International Journal of Aerospace Engineering*, 2020: e8816187, 2020. Publisher: Hindawi. 2

[3] Hoang Anh Dung, Bo Chen, and Tat-Jun Chin. A Spacecraft Dataset for Detection, Segmentation and Parts Recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2012–2019, Nashville, TN, USA, 2021. IEEE. 2

[4] Niccolò Faraco, Michele Maestrini, and Pierluigi Di Lizia. Instance Segmentation for Feature Recognition on Noncooperative Resident Space Objects. *Journal of Spacecraft and Rockets*, 59(6):2160–2174, 2022. 2

[5] Arianna Issitt, Trupti Mahendrakar, Amy Alvarez, Ryan T. White, and Alex Sizemore. On optimal observation orbits for learning gaussian splatting-based 3d models of unknown resident space objects. In *AIAA SCITECH 2025 Forum*, 2025. 1, 2, 6

[6] Arianna Issitt, Trupti Mahendrakar, and Ryan T. White. Reliable onboard 3d reconstruction of unknown spacecraft via data acquisition guided by orbital geometry and lighting. In *AIAA SCITECH 2026 Forum*, 2026. 1, 2

[7] Trupti Mahendrakar, Andrew Ekblad, Nathan Fischer, Ryan White, Markus Wilde, Brian Kish, and Isaac Silver. Performance Study of YOLOv5 and Faster R-CNN for Autonomous Navigation around Non-Cooperative Targets. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–12. IEEE, 2022. 2

[8] Trupti Mahendrakar, Ryan T. White, Madhur Tiwari, and Markus Wilde. Unknown non-cooperative spacecraft characterization with lightweight convolutional neural networks. *Journal of Aerospace Information Systems*, 21(5):455–460, 2024. 2, 5

[9] Trupti Mahendrakar, Ryan T. White, Madhur Tiwari, and Markus Wilde. Real-time, On-board Spacecraft Feature Detection with Lightweight CNNs. *AIAA Journal of Aerospace Information*, 2024. 2

[10] Mackenzie Meni, Trupti Mahendrakar, Olivia D. Raney, Ryan T. White, Michael L. Mayo, and Kevin R. Pilkiewicz. Taking a peek into yolov5 for satellite component recognition via entropy-based visual explanations. In *AIAA SCITECH 2024 Forum*. American Institute of Aeronautics and Astronautics, 2024. 2, 5, 7

[11] Mackenzie J. Meni. *Decoding Neural Networks: An Information-Theoretic Guide to Interpretability, Error Analysis and Efficiency*. Ph.d. dissertation, Florida Institute of Technology, Melbourne, Florida, 2024. 8

[12] Mackenzie J. Meni, Blake Gisclair, Ryan T. White, and Trupti Mahendrakar. Peek-guided neural network pruning for deployment on low swap hardware. In *39th Annual Small Satellite Conference*, 2025. 2

[13] Mackenzie J. Meni, Zigfried Hampel-Arias, and Adra Carr. Enhancing explainability in hyperspectral target identification using peek-based feature attribution. In *Applications of Machine Learning 2025*, page 2. SPIE, 2025. 7

[14] Mackenzie J. Meni, Trupti Mahendrakar, Olivia D. M. Raney, Ryan T. White, Michael L. Mayo, and Kevin R. Pilkiewicz. Probabilistic explanations for entropic knowledge extraction for automated satellite component detection. *Journal of Aerospace Information Systems*, 22(4):296–309, 2025. 2, 6

[15] NASA. Ice, Cloud, and Land Elevation Satellite (ICESat) 3D Model. https://science.nasa.gov/3d-resources/ice-clouds-and-land-elevation-satellite-icesat-b/, 2024. NASA 3D Resources. Accessed 2026. 3, 5

[16] NASA. Jason-1 3D Model. https://science.nasa.gov/3d-resources/jason-1/, 2024. NASA 3D Resources. Accessed 2026. 5

[17] NASA. Clementine 3D Model. https://science.nasa.gov/3d-resources/clementine/, 2025. NASA 3D Resources. Accessed 2026. 3, 5

[18] Tae Ha Park. Spacecraft Pose Estimation Dataset (SPEED). 2022. 2

[19] Tae Ha Park and Simone D'Amico. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. *Advances in Space Research*, 73(11): 5726–5740, 2024. 2

[20] Tae Ha Park and Simone D'Amico. Bridging the domain gap for flight-ready spaceborne vision. *Journal of Spacecraft and Rockets*, 63(1):289–307, 2026. 2

[21] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. SPEED+: Next-Generation Dataset for Spacecraft Pose Estimation across Domain Gap. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–15, 2022. ISSN: 1095-323X. 2

[22] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From coarse to fine: Robust hierarchical localization at large scale. In *CVPR*, 2019. 6

[23] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-Motion Revisited. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016. 6

[24] Sumant Sharma and Simone D'Amico. Neural Network-Based Pose Estimation for Noncooperative Spacecraft Rendezvous. *IEEE Transactions on Aerospace and Electronic Systems*, 56(6):4638–4658, 2020. 2

[25] Ultralytics. Ultralytics yolo. https://github.com/ultralytics/ultralytics, 2026. 4

[26] Szabolcs Velkei, Csaba Goldschmidt, and Károly Vass. A large-scale, physically-based synthetic dataset for satellite pose estimation. 2025. 2

[27] Herbert Viggh, Sean Loughran, Yaron Rachlin, Ross Allen, and Jessica Ruprecht. Training Deep Learning Spacecraft Component Detection Algorithms Using Synthetic Image Data. In *2023 IEEE Aerospace Conference*, pages 1–13, 2023. ISSN: 1095-323X. 2