Onboard Latent Kalman Filtering for Robust Spacecraft Pose Estimation

Anonymous CVPR submission

Paper ID 5

Abstract

Vision-based satellite pose estimation is critical for future space missions such as docking and debris removal. Although deep learning has advanced this field, the scarcity of real-world data forces reliance on synthetic imagery, leading to a persistent sim2real gap. Prior work mitigates this gap through computationally demanding online adaptation. In contrast, we address it without any online learning by leveraging temporal dynamics during rendezvous. We propose Dynamics-aware Robust Inference with latent Filtering for Temporal satellite pose estimation (DRIFT), a latent-space filtering framework that combines dynamics prediction with current observations. We first train a single-image encoder–regressor to extract pose latents, then freeze it and learn a lightweight Neural Kalman Gain module that produces adaptive filtering gains from innovations and recent latents. By explicitly modeling temporal dynamics in latent space, DRIFT enables temporally coherent and robust pose estimation in challenging sequential scenarios without onboard retraining.

1. Introduction

Vision-based satellite pose estimation is critical for on-orbit servicing, debris removal, and autonomous docking. The goal is to recover the 6-DoF pose of a target spacecraft from monocular images under challenging space conditions such as extreme illumination, weak texture, and limited onboard computation.

A major challenge is the scarcity of labeled real orbital imagery. Consequently, most methods rely on synthetic renderings for training. Recent learning-based models achieve strong single-frame accuracy [16], yet their deployment still suffers from a persistent simulation-to-reality (sim2real) gap [18]. Since acquiring real in-orbit image–pose pairs is extremely difficult, Hardware-in-the-Loop (HIL) evaluation is often used as a practical proxy [17]. To further reduce domain mismatch, some approaches adopt online or in-flight adaptation [26, 27]; however, such strategies are often impractical for on-orbit systems operating under strict computational and reliability constraints. This motivates lightweight temporal refinement mechanisms that improve robustness without retraining.

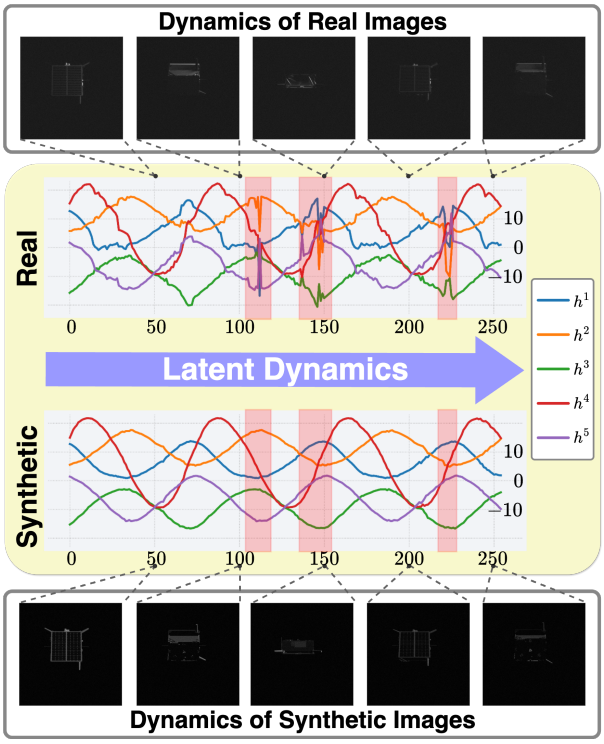


Figure 1. Latent dynamics over time for real (top) vs. synthetic (bottom) data. Each curve shows a latent component h^i over time. Smooth synthetic dynamics and spiky real dynamics reveal a sim2real gap that motivates our approach.

Indeed, as shown in Fig. 1, latent trajectories remain smooth in synthetic data but become unstable in real sequences, indicating that the sim2real gap also manifests in temporal dynamics. This observation motivates filtering directly in latent space to enforce temporal consistency while preserving the underlying perception model.

We propose DRIFT (Dynamics-aware Robust Inference with latent Filtering for Temporal pose estimation), a two-phase framework combining accurate single-image regres-

sion with learned latent-space filtering. First, we train an encoder–regressor for per-frame 6-DoF prediction. Then, operating in latent space, we extrapolate dynamics from a short history and fuse it with the current observation using an adaptive neural gain, yielding a filtered latent that is mapped to pose.

Our contributions are:

- **Latent dynamics filtering:** Sequential pose estimation formulated as Kalman-style filtering in feature space with a robust velocity prior.
- **Neural Kalman Gain:** A lightweight MLP that adaptively fuses observation and dynamics without hand-tuned noise models.
- **Improved sim2real robustness:** Enhanced temporal stability and accuracy under domain shift and degraded observations without online retraining.

2. Related Work

2.1. Satellite Pose Estimation

Monocular satellite pose estimation has traditionally relied on detecting 2D features and solving a Perspective-*n*-Point (PnP) problem for 6-DoF recovery [9, 24]. Learning-based extensions predict sparse landmarks or heatmaps followed by PnP, improving robustness while preserving geometric structure [19]. More recent end-to-end CNN/ViT models directly regress pose or intermediate representations, sometimes augmented with lightweight temporal refiners or classical filters [13, 14].

Due to the scarcity of labeled orbital imagery, most methods are trained on synthetic data and suffer from a persistent sim2real gap. The SPEED+ benchmark [17], built on synthetic and Hardware-in-the-Loop data from TRON [23], formalized this issue, and subsequent work still reports domain mismatch [5, 21]. To mitigate this gap, several methods perform onboard adaptation via UKF-based pseudo-labeling, self-training, or motion-consistency optimization [10, 15, 26, 27], albeit at increased computational and operational complexity.

2.2. Temporal Filtering and Learned Kalman Gains

Classical Kalman filtering (KF) [6] provides a recursive predictor–corrector framework but depends on accurate dynamics and noise modeling. Recent works therefore learn components of KF while retaining its update structure. Neural–KF hybrids such as BKF and LSTM-KF [2, 4] integrate learned observation or motion models, while deep generative state-space models (e.g., Deep KF, DVBF, KVAE) perform latent filtering via variational inference [3, 7, 8]. More recent methods directly learn the Kalman gain, including RKN, KalmanNet, and KalmanFormer [1, 22, 25].

Unlike these approaches, which jointly learn dynamics and filtering in explicitly modeled state spaces, we adopt representation-decoupled latent filtering: the pre-trained pose encoder is frozen, and only a lightweight element-wise neural gain is learned, enabling plug-and-play temporal refinement without explicit covariance modeling.

3. Proposed Method: DRIFT

We address sequential satellite pose estimation from an image stream $\{\mathcal{X}_t\}_{t=0}^{T}$ by formulating the problem in a latent state–space framework. At each time step t, we denote the filtered latent state by $\mathbf{h}_t \in \mathbb{R}^d$, and the observation latent extracted from the current image by

$$\mathbf{h}_t^{obs} = \psi(\mathcal{X}_t), \tag{1}$$

where ψ is an image encoder.

3.1. Latent State-Space Formulation

The latent evolution is modeled as a nonlinear stochastic system:

$$\mathbf{h}_t^{dyn} = \mathcal{A}(\mathbf{h}_{t-1}) + \mathbf{w}_{t-1}, \qquad \mathbf{w}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t), \tag{2}$$

$$\mathbf{h}_t = \mathcal{B}(\mathbf{h}_t^{dyn}, \mathbf{h}_t^{obs}) + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t), \tag{3}$$

where $\mathbf{h}_t^{dyn}$ is the dynamics-based prediction, $\mathbf{w}_{t-1}$ and $\mathbf{v}_t$ represent process and measurement noise with covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$, respectively, and $\mathcal{B}(\cdot,\cdot)$ denotes a generic fusion operator. We initialize $\mathbf{h}_0 = \psi(\mathcal{X}_0)$.

This formulation parallels the classical linear–Gaussian Kalman filter:

$$\mathbf{h}_t = \mathbf{F}_t \mathbf{h}_{t-1} + \mathbf{w}_{t-1}, \qquad \mathbf{h}_t^{obs} = \mathbf{H}_t \mathbf{h}_t + \mathbf{v}_t, \tag{4}$$

which alternates between prediction and update. The prediction step propagates the state estimate and covariance:

$$\hat{\mathbf{h}}_{t|t-1} = \mathbf{F}_t\, \hat{\mathbf{h}}_{t-1|t-1}, \tag{5}$$

$$\mathbf{P}_{t|t-1} = \mathbf{F}_t\, \mathbf{P}_{t-1|t-1}\, \mathbf{F}_t^{\top} + \mathbf{Q}_t, \tag{6}$$

where $\mathbf{F}_t \in \mathbb{R}^{d\times d}$ is the state-transition matrix, $\mathbf{P}_{t|t-1} \in \mathbb{R}^{d\times d}$ is the predicted state-error covariance, $\mathbf{Q}_t \in \mathbb{R}^{d\times d}$ is the process-noise covariance, and $\mathbf{H}_t \in \mathbb{R}^{d\times d}$ is the measurement matrix. The update step based on the innovation $\boldsymbol{\nu}_t$ and its covariance $\mathbf{S}_t$ can be expressed as follows:

$$\boldsymbol{\nu}_t = \mathbf{h}_t^{obs} - \mathbf{H}_t\, \hat{\mathbf{h}}_{t|t-1}, \quad \mathbf{S}_t = \mathbf{H}_t\, \mathbf{P}_{t|t-1}\, \mathbf{H}_t^{\top} + \mathbf{R}_t, \tag{7}$$

with Kalman gain $\mathbf{K}_t$ and filtered state $\mathbf{h}_t$:

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\, \mathbf{H}_t^{\top}\, \mathbf{S}_t^{-1}, \qquad \mathbf{h}_t = \hat{\mathbf{h}}_{t|t-1} + \mathbf{K}_t\, \boldsymbol{\nu}_t. \tag{8}$$

In our latent setting, we assume an identity observation mapping ($\mathbf{H}_t = \mathbf{I}$) and replace $\hat{\mathbf{h}}_{t|t-1}$ with $\mathbf{h}_t^{dyn}$. The innovation therefore reduces to

$$\boldsymbol{\nu}_t = \mathbf{h}_t^{obs} - \mathbf{h}_t^{dyn}. \tag{9}$$

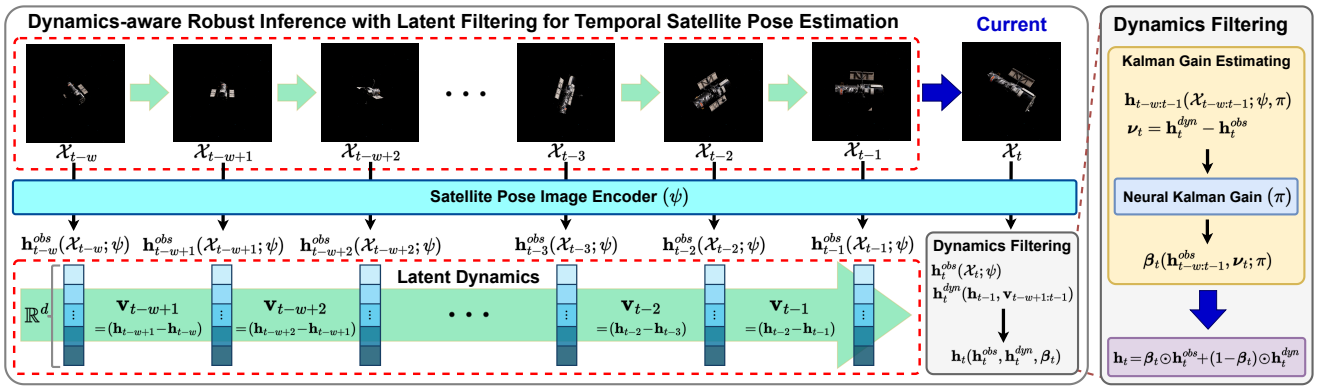


Figure 2. Latent dynamics filtering: Given a short history of filtered latents $\mathbf{h}_{t-w:t-1}$, we form a dynamics prediction $\mathbf{h}_t^{dyn}$, compute the innovation $\boldsymbol{\nu}_t$, and use the NKG to produce a parameterized Kalman gain $\boldsymbol{\beta}_t$. The current filtered state is $\mathbf{h}_t$, which is mapped by the frozen regressor ϕ to the final pose $\hat{E}$.

When the gain is restricted to a diagonal form, $\mathbf{K}_t = \mathrm{diag}(\boldsymbol{\beta}_t)$ with $\boldsymbol{\beta}_t \in [0,1]^d$, the update becomes an element-wise convex combination:

$$\mathbf{h}_t = (1-\boldsymbol{\beta}_t) \odot \mathbf{h}_t^{dyn} + \boldsymbol{\beta}_t \odot \mathbf{h}_t^{obs}. \tag{10}$$

In classical KF, $\mathbf{K}_t$ depends explicitly on the propagated covariances $\mathbf{P}_{t|t-1}$, $\mathbf{Q}_t$, and $\mathbf{R}_t$, requiring accurate system modeling and careful noise tuning. In contrast, DRIFT avoids explicit covariance propagation and instead learns $\boldsymbol{\beta}_t$ directly in latent space, while preserving the principled prediction–innovation–update structure of Kalman filtering.

3.2. Phase 1: Single-Image Pose Learning

Phase 1 trains an encoder–regressor pair to estimate pose from a single image. The encoder ψ maps $\mathcal{X}_t$ to a compact latent $\mathbf{h}_t^{obs} \in \mathbb{R}^d$, and the regressor ϕ predicts 6-DoF pose

$$\hat{E}_t = \phi(\mathbf{h}_t^{obs}), \quad \hat{E}_t^{tran} \in \mathbb{R}^3, \quad \hat{E}_t^{rot} \in SO(3). \tag{11}$$

We adopt a geodesic rotation loss and normalized translation error:

$$\mathcal{L} = \arccos\Big(\frac{\mathrm{Tr}(\hat{E}^{rot} E^{rot\top}) - 1}{2}\Big) + \frac{\|\hat{E}^{tran} - E^{tran}\|_2}{\|E^{tran}\|_2}. \tag{12}$$

After convergence, both ψ and ϕ are frozen.

3.3. Phase 2: Dynamics-Aware Latent Filtering

With ψ and ϕ fixed, we learn a Neural Kalman Gain (NKG) that adaptively determines $\boldsymbol{\beta}_t$ in Eq. (10).

Robust latent dynamics. Given a filtered history $\mathbf{h}_{t-w:t-1}$, we compute velocity increments

$$\mathbf{v}_i = \mathbf{h}_i - \mathbf{h}_{i-1}, \quad i = t-w+1, \ldots, t-1, \tag{13}$$

and estimate dynamics via median extrapolation:

$$\mathbf{h}_t^{dyn} = \mathbf{h}_{t-1} + \mathrm{median}(\mathbf{v}_{t-w+1:t-1}). \tag{14}$$

Using filtered latents prevents error accumulation, while the median operator suppresses outliers.

Neural Kalman Gain. Instead of deriving $\mathbf{K}_t$ from covariances, we learn an element-wise gain:

$$\boldsymbol{\beta}_t = \mathcal{K}_\pi\big(\boldsymbol{\nu}_t, \mathbf{h}_{t-w:t-1}\big), \tag{15}$$

where $\mathcal{K}_\pi$ is a lightweight MLP. The filtered latent $\mathbf{h}_t$ is then mapped to pose via $\hat{E}_t = \phi(\mathbf{h}_t)$.

Noise-Robust Training. Because the NKG is trained mainly on synthetic sequences with overly clean latents, directly optimizing Eq. (12) can lead to a trivial solution that overly trusts $\mathbf{h}_t^{obs}$. To promote adaptive gain learning and improve sim-to-real robustness, we inject stochastic corruption into the observation latent during training:

$$\tilde{\mathbf{h}}_t^{obs} = \mathbf{h}_t^{obs} + \mathbf{m}_t \odot \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}), \tag{16}$$

where $\mathbf{m}_t$ is a Bernoulli mask with probability p_{noise}.

4. Experiments

We evaluate DRIFT on public sequential spacecraft pose benchmarks and our HST sequences, comparing against strong baselines in terms of per-frame accuracy and sequence-level dynamics. We report results on SHIRT and HST (ROE1/ROE2) and analyze latent trajectory behavior.

4.1. Experimental Setup

Datasets We use SPEED+ [17] to pre-train a single-frame pose encoder and regressor on synthetic OpenGL images only; these models are frozen thereafter.

Table 1. Quantitative comparison on **SHIRT** and **HST** (ROE 1/ROE 2). Lower is better; the best results are shown in **bold**.

Dataset	Metric	SPNv2	SPNv3	OST-v2	OST-v3	Ours
SHIRT (ROE 1)	$\mathcal{E}_{\text{tran}}$ [m]	0.196	0.213	12.0	11.0	**0.195**
	$\mathcal{E}_{\text{rot}}$ [°]	11.0	9.16	22.8	13.4	**6.75**
	$\mathcal{E}_{\text{p}}$ [-]	0.232	0.203	2.80	2.44	**0.157**
SHIRT (ROE 2)	$\mathcal{E}_{\text{tran}}$ [m]	0.178	**0.135**	8.03	8.21	0.177
	$\mathcal{E}_{\text{rot}}$ [°]	5.90	6.32	15.2	13.5	**5.64**
	$\mathcal{E}_{\text{p}}$ [-]	0.139	0.137	1.87	1.88	**0.134**
HST (ROE 1)	$\mathcal{E}_{\text{tran}}$ [m]	0.488	0.524	0.570	0.521	**0.344**
	$\mathcal{E}_{\text{rot}}$ [°]	4.32	9.53	7.72	9.60	**3.23**
	$\mathcal{E}_{\text{p}}$ [-]	0.0997	0.193	0.163	0.194	**0.0736**
HST (ROE 2)	$\mathcal{E}_{\text{tran}}$ [m]	0.939	**0.889**	1.07	0.912	0.930
	$\mathcal{E}_{\text{rot}}$ [°]	2.11	5.18	4.06	5.14	**1.97**
	$\mathcal{E}_{\text{p}}$ [-]	0.0838	0.135	0.125	0.135	**0.0809**

For temporal modeling, we use the sequential SHIRT benchmark [14], which provides synthetic and TRON-captured HIL trajectories under two rendezvous scenarios (ROE1: single-axis rotation; ROE2: tumbling with approach). The synthetic split is used to train the neural gain module, and the HIL split is used for sim2real evaluation.

To assess target generalization, we additionally construct a synthetic sequential HST dataset based on the public NASA CAD model, following SHIRT-like trajectories.

Evaluation Metrics We report translation error $\mathcal{E}_{\text{tran}}$, rotation error $\mathcal{E}_{\text{rot}}$, and combined pose error $\mathcal{E}_p$ following prior works [11, 12, 16, 20]:

$$\mathcal{E}_{\text{tran}} = \|\hat{E}^{tran} - E^{tran}\|_2, \tag{17}$$

$$\mathcal{E}_{\text{rot}} = 2\arccos\left(\frac{\text{Tr}(\hat{E}^{rot}E^{rot\top}) - 1}{2}\right), \tag{18}$$

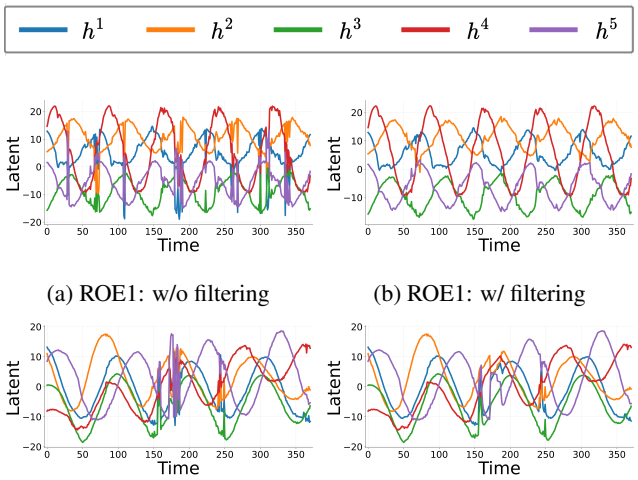
$$\mathcal{E}_p = \mathcal{E}_{\text{rot}} + \frac{\mathcal{E}_{\text{tran}}}{\|E^{tran}\|_2}. \tag{19}$$

4.2. Baselines

We compare DRIFT with representative satellite pose estimators: SPNv2 [16] (CNN-based), SPNv3 [13] (transformer-based), and OST [15], an online adaptation method built on these backbones (OST-v2, OST-v3).

4.3. Quantitative Results on SHIRT and HST

Table 1 reports rotation and translation errors on SHIRT–Lightbox and HST under ROE1/ROE2. DRIFT consistently outperforms all baselines, reducing rotation and combined pose errors across datasets while maintaining competitive or best translation accuracy. Gains are most pronounced on ROE1 and remain stable on ROE2. In contrast, OST exhibits unstable updates in limited-view scenarios such as ROE1, leading to degraded robustness. DRIFT achieves consistent improvements without any onboard retraining, demonstrating that latent-space filtering effectively captures motion regularities for sim2real generalization.



(a) ROE1: w/o filtering (b) ROE1: w/ filtering

(c) ROE2: w/o filtering (d) ROE2: w/ filtering

Figure 3. Latent trajectories on SHIRT–Real for ROE1 and ROE2, shown without (a,c) and with (b,d) filtering. Each curve denotes a latent component h^i; the legend (h^1–h^5) is shown above.

4.4. Latent Trajectory Analysis

To probe the mechanism behind these gains, Fig. 3 visualizes the trajectories of individual latent components $\mathbf{h} = \{h^1, \ldots, h^r\}$ with and without latent filtering. Unfiltered latents show high-frequency oscillations and outliers, whereas DRIFT produces smoother, more consistent temporal patterns, particularly in ROE1. This behavior reflects innovation-aware gain modulation, which suppresses unreliable observations and stabilizes latent evolution. These results suggest that latent-space filtering improves robustness by enforcing temporal coherence without any online or onboard retraining.

5. Conclusions

We presented DRIFT, a computationally efficient framework for sequential satellite pose estimation that decouples single-frame perception from latent-space temporal filtering, eliminating the need for online adaptation. By combining a robust dynamics prior with a learned neural gain, DRIFT produces stable and temporally coherent pose estimates, consistently improving over strong single-frame baselines. Its lightweight design introduces minimal overhead and enables practical, deployable onboard sim2real-robust perception for on-orbit servicing and docking. More broadly, DRIFT highlights the effectiveness of latent-space temporal modeling as a principled alternative to in-flight adaptation for resource-constrained onboard space systems.

References

[1] Philipp Becker, Harit Pandya, Gregor Gebhardt, Cheng Zhao, C James Taylor, and Gerhard Neumann. Recurrent kalman networks: Factorized inference in high-dimensional deep feature spaces. In *International conference on machine learning*, pages 544–552. PMLR, 2019. 2

[2] Huseyin Coskun, Felix Achilles, Robert DiPietro, Nassir Navab, and Federico Tombari. Long short-term memory kalman filters: Recurrent neural estimators for pose regularization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5524–5532, 2017. 2

[3] Marco Fraccaro, Simon Kamronn, Ulrich Paquet, and Ole Winther. A disentangled recognition and nonlinear dynamics model for unsupervised learning. *Advances in neural information processing systems*, 30, 2017. 2

[4] Tuomas Haarnoja, Anurag Ajay, Sergey Levine, and Pieter Abbeel. Backprop kf: Learning discriminative deterministic state estimators. *Advances in neural information processing systems*, 29, 2016. 2

[5] Mohsi Jawaid, Ethan Elms, Yasir Latif, and Tat-Jun Chin. Towards bridging the space domain gap for satellite pose estimation using event sensing. *arXiv preprint arXiv:2209.11945*, 2022. 2

[6] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. 1960. 2

[7] Maximilian Karl, Maximilian Soelch, Justin Bayer, and Patrick Van der Smagt. Deep variational bayes filters: Unsupervised learning of state space models from raw data. *arXiv preprint arXiv:1605.06432*, 2016. 2

[8] Rahul G Krishnan, Uri Shalit, and David Sontag. Deep kalman filters. *arXiv preprint arXiv:1511.05121*, 2015. 2

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

[10] Kun Liu and Yongjun Yu. Revisiting the domain gap issue in non-cooperative spacecraft pose tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6864–6873, 2024. 2

[11] Alessandro Lotti, Dario Modenini, Paolo Tortora, Massimiliano Saponara, and Maria A Perino. Deep learning for real time satellite pose estimation on low power edge tpu. *arXiv preprint arXiv:2204.03296*, 2022. 4

[12] Marcus Märtens and Simone D'Amico. Satellite pose estimation challenge: Dataset, competition design and results. 4

[13] Tae Ha Park and Simone D'Amico. Bridging domain gap for flight-ready spaceborne vision. *arXiv preprint arXiv:2409.11661*, 2024. 2, 4

[14] Tae Ha Park and Simone D'Amico. Adaptive neural-network-based unscented kalman filter for robust pose tracking of noncooperative spacecraft. *Journal of Guidance, Control, and Dynamics*, 46(9):1671–1688, 2023. 2, 4

[15] Tae Ha Park and Simone D'Amico. Online supervised training of spaceborne vision during proximity operations using adaptive kalman filtering. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11744–11752. IEEE, 2024. 2, 4

[16] 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. 1, 4

[17] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In *2022 IEEE aerospace conference (AERO)*, pages 1–15. IEEE, 2022. 1, 2, 3

[18] Leo Pauly, Wassim Rharbaoui, Carl Shneider, Arunkumar Rathinam, Vincent Gaudilliere, and Djamila Aouada. A survey on deep learning-based monocular spacecraft pose estimation: Current state, limitations and prospects. *Acta Astronautica*, 212:339–360, 2023. 1

[19] Juan Ignacio Bravo Pérez-Villar, Álvaro García-Martín, and Jesús Bescós. Spacecraft pose estimation based on unsupervised domain adaptation and on a 3d-guided loss combination. In *European Conference on Computer Vision*, pages 37–52. Springer, 2022. 2

[20] Pedro F Proença and Yang Gao. Deep learning for spacecraft pose estimation from photorealistic rendering. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6007–6013. IEEE, 2020. 4

[21] Arunkumar Rathinam, Haytam Qadadri, and Djamila Aouada. Spades: A realistic spacecraft pose estimation dataset using event sensing. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11760–11766. IEEE, 2024. 2

[22] Guy Revach, Nir Shlezinger, Xiaoyong Ni, Adria Lopez Escoriza, Ruud JG Van Sloun, and Yonina C Eldar. Kalmannet: Neural network aided kalman filtering for partially known dynamics. *IEEE Transactions on Signal Processing*, 70: 1532–1547, 2022. 2

[23] Sumant Sharma, Tae Ha Park, and Simone D'Amico. Spacecraft pose estimation dataset (speed). *Stanford Digital Repository*, 2019. 2

[24] Sumant Sharma et al. Comparative assessment of techniques for initial pose estimation using monocular vision. *Acta Astronautica*, 123:435–445, 2016. 2

[25] Siyuan Shen, Jichen Chen, Guanfeng Yu, Zhengjun Zhai, and Pujie Han. Kalmanformer: using transformer to model the kalman gain in kalman filters. *Frontiers in Neurorobotics*, 18:1460255, 2025. 2

[26] Zi Wang, Minglin Chen, Yulan Guo, Zhang Li, and Qifeng Yu. Bridging the domain gap in satellite pose estimation: A self-training approach based on geometrical constraints. *IEEE transactions on aerospace and electronic systems*, 60 (3):2500–2514, 2023. 1, 2

[27] Yunfeng Zhang, Linjing You, Luyu Yang, Zhiwei Zhang, Xiangli Nie, and Bo Zhang. Monocular 6-dof pose estimation of spacecrafts utilizing self-iterative optimization and motion consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6847–6856, 2024. 1, 2