Leveraging neural network uncertainty in adaptive unscented Kalman Filter for spacecraft pose estimation

Cassinis, Lorenzo Pasqualetto; Park, Tae Ha; Stacey, Nathan; D'Amico, Simone; Menicucci, Alessandra; Gill, Eberhard; Ahrns, Ingo; Sanchez-Gestido, Manuel

DOI
10.1016/j.asr.2023.02.021

Publication date
2023

Document Version
Final published version

Published in
Advances in Space Research

Citation (APA)
Cassinis, L. P., Park, T. H., Stacey, N., D'Amico, S., Menicucci, A., Gill, E., Ahrns, I., & Sanchez-Gestido, M. (2023). Leveraging neural network uncertainty in adaptive unscented Kalman Filter for spacecraft pose estimation. *Advances in Space Research*, *71*(12), 5061-5082. https://doi.org/10.1016/j.asr.2023.02.021

Available online at www.sciencedirect.com

ScienceDirect

Advances in Space Research 71 (2023) 5061–5082

ADVANCES IN SPACE RESEARCH (a COSPAR publication)

www.elsevier.com/locate/asr

Leveraging neural network uncertainty in adaptive unscented Kalman Filter for spacecraft pose estimation

Lorenzo Pasqualetto Cassinis [a,*], Tae Ha Park [b], Nathan Stacey [b], Simone D'Amico [b], Alessandra Menicucci [a], Eberhard Gill [a], Ingo Ahrns [c], Manuel Sanchez-Gestido [d]

[a] Delft University of Technology, Kluyverweg 1 2629 HS, Delft, The Netherlands
[b] Stanford University, 496 Lomita Mall, 94305 Palo Alto, CA, United States
[c] Airbus DS GmbH, Airbusallee 1, 28199 Bremen, Germany
[d] ESTEC, Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands

Received 29 June 2022; received in revised form 6 January 2023; accepted 12 February 2023
Available online 18 March 2023

Abstract

This paper introduces an adaptive Convolutional Neural Network (CNN)-based Unscented Kalman Filter for the pose estimation of uncooperative spacecraft. The validation is carried out at Stanford's robotic Testbed for Rendezvous and Optical Navigation on the Satellite Hardware-In-the-loop Rendezvous Trajectories (SHIRT) dataset, which simulates vision-based rendezvous trajectories of a servicer spacecraft to PRISMA's Tango spacecraft. The proposed navigation system is stress-tested on synthetic as well as realistic lab imagery by simulating space-like illumination conditions on-ground. The validation is performed at different levels of the navigation system by first training and testing the adopted CNN on SPEED+, Stanford's spacecraft pose estimation dataset with specific emphasis on domain shift between a synthetic domain and an Hardware-In-the-Loop domain. A novel data augmentation scheme based on light randomization is proposed to improve the CNN robustness under adverse viewing conditions, reaching centimeter-level and 10 degree-level pose errors in 80% of the SPEED+ lab images. Next, the entire navigation system is tested on the SHIRT dataset. Results indicate that the inclusion of a new scheme to adaptively scale the heatmaps-based measurement error covariance based on filter innovations improves filter robustness by returning centimeter-level position errors and moderate attitude accuracies, suggesting that a proper representation of the measurements uncertainty combined with an adaptive measurement error covariance is key in improving the navigation robustness.

Keywords: Relative pose estimation; Active debris removal; Adaptive filtering; On-ground validation; Convolutional neural networks; Domain adaptation

2010 MSC: 00-01; 99-00

1. Introduction

Nowadays, more than 30,000 monitored pieces of debris are orbiting the Earth, including non-functional spacecraft, abandoned launch vehicle stages, and fragmentation debris[1]. Altogether, this debris pose a threat to commercial

* Corresponding author.
E-mail addresses: L.PasqualettoCassinis@tudelft.nl (L.P. Cassinis), tpark94@stanford.edu (T.H. Park), nstacey@stanford.edu (N. Stacey), damicos@stanford.edu (S. D'Amico), A.Menicucci@tudelft.nl (A. Menicucci), E.K.A.Gill@tudelft.nl (E. Gill), ingo.ahrns@airbus.com (I. Ahrns), Manuel.Sanchez.Gestido@esa.int (M. Sanchez-Gestido).

https://doi.org/10.1016/j.asr.2023.02.021
0273-1177/

[1] https://www.esa.int/Safety_Security/Space_Debris/Space_debris_by_the_numbers.

satellites as well as human spaceflight, undermining safety and operations in orbit. At the same time, autonomous refueling and repairing is regarded as the most viable solution to extend the lifetime of active satellites in orbit. In this context, advancements in the field of autonomous navigation were made in the past years to enable key technologies for future In-Orbit Servicing (IOS) and Active Debris Removal (ADR) missions (Tatsch et al., 2006; Wieser et al., 2015). For such scenarios, the estimation of the pose (relative position and attitude) of an uncooperative target object by an active servicer spacecraft represents a critical navigation task. Compared to cooperative close-proximity missions, the pose estimation problem is complicated by the fact that the target object does not aid the relative navigation and may not be functional. Hence, optical sensors on the servicer spacecraft are well suited to pose estimation because they do not rely on any sensing hardware on the target, such as Global Navigation Satellite Systems (GNSS) sensors and antennas. In this context, pose estimation systems based solely on a monocular camera are recently becoming an attractive alternative to systems based on active sensors or stereo cameras, due to their reduced mass, power consumption and system complexity (Sharma et al., 2018; Pasqualetto Cassinis et al., 2019). However, a significant effort is still required to comply with most of the demanding requirements for a robust and accurate monocular-based relative navigation system. Notably, the aforementioned navigation system cannot rely on known visual markers, as they are typically not available on an uncooperative target. Since the extraction of visual features is an essential step in the pose estimation process, advanced Image Processing (IP) techniques are required to extract keypoints (or interest points), corners, and edges on the target body, such as the Speeded Up Robust Features (SURF) Bay et al. (2008), the Scale Invariant Feature Transform (SIFT) (Lowe, 2004), and the Oriented FAST, Rotated BRIEF (ORB) Rublee et al. (2011). In model-based methods, the detected features are then matched with pre-defined features on an offline wireframe 3D model of the target to solve for the pose. This is usually achieved by solving the Perspective-n-Points (PnP) problem Sharma and D'Amico (2015) with standard pose estimation solvers such as the Efficient Perspective-n-Point (EPnP) Lepetit and Moreno-Noguer (2009), the Efficient Procrustes Perspective-n-Point (EPPnP) Ferraz et al. (2014b), or the multi-dimensional Newton Raphson Method (NRM) Ostrowsky (1966), additionally using the RANdom SAmple Consensus (RANSAC) method to improve the robustness against outliers. Essentially, a reliable detection of key features is critical to guarantee safe operations around an uncooperative target, e.g. under adverse illumination conditions.

Unfortunately, standard IP algorithms usually lack feature detection robustness when applied to space images D'Amico et al. (2014), undermining the overall navigation system and, in turn, the close-proximity operations around the uncooperative target. From a pose initialization standpoint, the extraction of target features can in fact be jeopardized by external factors, such as adverse illumination conditions, low Signal-to-Noise ratio (SNR) and Earth in the background, as well as by target-specific factors, such as the presence of complex textures and features on the target body. Moreover, most of the IP methods are based on the image gradient, detecting textured-rich features or highly visible parts of the target silhouette Pasqualetto Cassinis et al. (2019). As such, the detected features are image-specific and can vary in number and typology depending on the image histogram. This means that most of these techniques cannot accommodate an offline feature selection step, resulting in a computationally expensive image-to-model correspondence process to ensure that each detected 2D feature is matched with its 3D counterpart on the available wireframe model of the target object.

In recent years, Convolutional Neural Networks (CNNs) are emerging as a valid alternative to more traditional pose estimation systems, with two main CNN-based architectures currently standing out. Initially, *end-to-end* architectures in which a single CNN replaced the entire pose estimation pipeline were adopted Sharma et al. (2018), Sharma and D'Amico (2020), Shi et al. (2018), Sonawani et al. (2020). However, since the pose accuracies of these systems proved to be lower than the accuracies returned by standard PnP solvers, especially in the estimation of the relative attitude Sharma et al. (2018), *keypoints-based* architectures stood out as the preferred option. Specifically, degree-level average orientation errors were achieved by keypoints-based methods as opposed to average orientation errors one order of magnitude larger achieved by end-to-end methods. These averages were computed across test images of the Tango spacecraft from the PRISMA mission D'Amico et al. (2013) as part of the Spacecraft Pose Estimation Dataset (SPEED) challenge Kisantal et al. (2020), Huo et al. (2020) (Sharma et al., 2019). In keypoints-based CNN systems, a CNN is used only at a feature detection level to replace standard IP algorithms, and the output features are fed to a PnP solver together with their body coordinates, which are made available through the wireframe 3D model of the target body. Due to the fact that the trainable features can be selected offline prior to the training, the matching of the extracted feature points with the features of the wireframe model can be performed without the need of a large search space for the image-model correspondences, which usually characterizes most of the edges/corners-based methods (D'Amico et al., 2014). However, due to a lack of availability of representative space images, these CNN systems often need to be trained with synthetic renderings of the available target model. As a result, their feature detection robustness on more realistic images is usually unknown and difficult to predict. Therefore, the synthetic datasets used to train the CNNs tend to fail in representing the textures of the target as well as the external illuminations, resulting in inaccurate detections and low pose estimation accuracies when tested on images from

different sources, such as spaceborne images (Kisantal et al., 2020; Pasqualetto Cassinis et al., 2022). In this context, three desirable aspects stand out. First of all, novel techniques shall be investigated to improve the performance of CNN-based pose estimation systems trained using synthetic images on actual space imagery. This aspect is referred to as the *domain shift problem* (Tobin et al., 2017; Ben-David et al., 2007; Sugiyama and Muller, 2005). Second, the interface between the CNN and the navigation filter shall be explored to assess the performance of a CNN-based navigation filter on representative relative trajectories. Third, a proper validation framework shall be sought to test the robustness of the proposed CNN-based system against representative images of the target spacecraft, generated in a laboratory environment which simulates space-like conditions. Notably, a calibration framework shall be established which returns an accurate reference for the pose between the monocular camera and the target mockup for each generated image (pose labels), in order to be able to quantify the CNN performance at both keypoints detection and pose estimation levels.

The domain shift problem in CNNs has been extensively investigated in recent years to leverage the domain gap from synthetic training to real test imagery, either via *data augmentation* (Tobin et al., 2017; Jackson et al., 2018; Geirhos et al., 2019) or via *domain adaptation* (Donahue et al., 2017, 2016). In relation to the latter, domain adaptation techniques were leveraged by the winners of the recent Spacecraft Pose Estimation Challenge (SPEC) to tackle the domain gap challenges in the Next Generation Satellite Pose Estimation Dataset (SPEED+) (Park et al., 2023). Although domain adaptation techniques are often effective and can produce impressive results by adapting the CNN on a specific target domain post training, they require the target domain images and synthetic training images simultaneously to perform adaptation, and hence they are not domain-agnostic. On the other hand, data augmentation techniques consist of introducing variations in the synthetic training domain without any a priori knowledge of the target domain. In essence, the idea is to extend the standard data augmentation effects, such as random cropping, zooming, rotation, flipping etc. with texture and complex illumination variations. By doing that, Tobin et al. (2017) showed that a CNN can generalize from synthetic environments to new domains by using an unrealistic but diverse set of random textures. Following this line of reasoning, Jackson et al. (2018) and Geirhos et al. (2019) further discovered that by randomizing textures during training, CNNs can learn the shape of objects rather than textures, improving their robustness to domain shift. By extending the above methods from terrestrial applications to space scenarios, some of the authors (Pasqualetto Cassinis et al., 2022; Park et al., 2021a; Park et al., 2019) already highlighted that augmenting the CNN training with texture randomization can lead to a considerable improvement in the overall pose estimation performance under large domain gaps in the target's texture. However, preliminary results showed that the CNN performance decreases under highly adverse illumination conditions, suggesting that additional augmentation techniques are required to tackle the domain shift from an illumination standpoint (Pasqualetto Cassinis et al., 2022).

Besides domain shift, additional challenges arise in relation to the applicability of CNNs for relative navigation in space. Notably, both pose initialization and pose tracking are not well suited to produce pose estimates at high frequencies, especially due to the computationally expensive IP in combination with pose estimation. Furthermore, solving for the pose solely from CNN detections can only provide a prediction from actual sensor data without accounting for any modeling of the external environment. Therefore, it is not guaranteed that the estimation can reliably deal with unwanted components in the measurements, i.e. image noise or adverse illumination conditions. Finally, quantities such as the translational and rotational velocities of the target spacecraft with respect to the servicer spacecraft can hardly be estimated together with the pose. Therefore, filtering techniques are usually used in combination with the camera measurements and the actual pose estimate in order to return full-state (pose and relative velocities) solutions at high frequency (Sharma and D'Amico, 2017). Many navigation filters for close-proximity operations were investigated in recent years in the context of pose estimation. The reader is referred to Pasqualetto Cassinis et al. (2019) for a comprehensive overview. However, limited focus has been given to the interface between a CNN-based system and a navigation filter. Earlier works by some of the authors (Pasqualetto Cassinis et al., 2021) suggested that a tightly-coupled approach should be more suited than a loosely-coupled one, if a CNN is used to extract features from the monocular images. As opposed to a loosely-coupled approach in which the detected features are transformed into an estimated pose prior to the filter, a tightly-coupled approach directly feeds the detected target features into the filter. However, limited research was provided on how to model the features detection uncertainty into a representative measurement error covariance. Other authors proposed a method based on projection vector, in which the covariance is associated to the image gradient magnitude and direction at each feature location (Cui et al., 2019), or a method in which covariance information is derived for each feature based on feature's visibility and robustness against illumination changes (Harvard et al., 2020). However, in all these methods the derivation of features covariance matrices is a lengthy process which generally cannot be directly related to the actual detection uncertainty. This is because it may not be possible to associate the gradient information to the image processing algorithm. Moreover, this procedure could not be easily applied if CNNs are used in the feature detection step, due to the difficulty to associate statistical meaning to the IP tasks performed within the network. In this context, a novel method was recently presented by some of the authors in Pasqualetto Cassinis et al. (2021), in which the

output of the CNNs is directly exploited to return relevant statistical information about the detection step. This method consists in creating a statistical distribution from the CNN heatmaps, and to derive a representative covariance matrix from each feature based on the heatmap's dispersion. However, despite the promising preliminary finding this heatmaps-based method has not been extensively tested on realistic space imagery. Specifically, it is still unclear whether the magnitude of such a covariance can effectively represent the feature detection uncertainty.

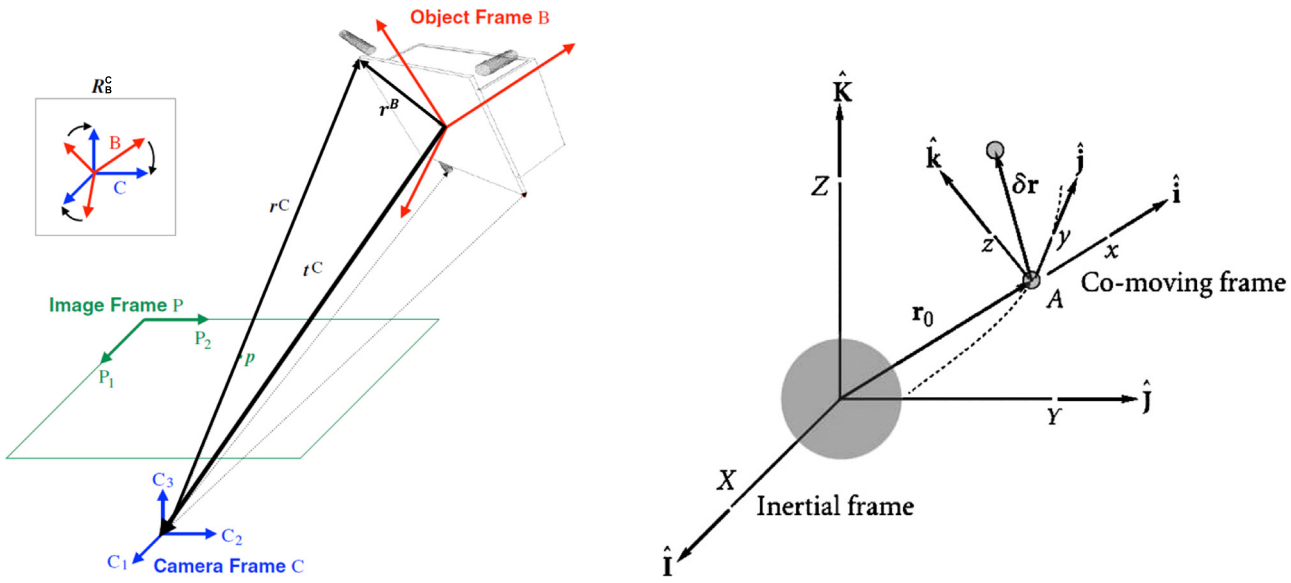
In the context of on-ground validation, several laboratory testbeds exist to generate Hardware-In-theLoop (HIL) images of a target spacecraft's mockup with a monocular camera (Wilde et al., 2019), e.g. the robotic Testbed for Rendezvous and Optical Navigation (TRON) at Stanford University (Kisantal et al., 2020), the GNC Rendezvous, Approach and Landing Simulator (GRALS) at the European Space Research and Technology Centre (ESTEC) (Zwick et al., 2018), the European Proximity Operations Simulator (EPOS) at the German Aerospace Agency (DLR) (Krüger and Theil, 2010), and the Platform-art facility at GMV (Dubanchet et al., 2020). The TRON testbed was the first of its kind to provide the generation of large-scale pose estimation datasets of a target mockup satellite under multiple camera-target geometries and illumination conditions. The resulting Spacecraft Pose Estimation Dataset (SPEED) and its extension SPEED+ guarantee a reliable validation of CNN-based pose estimation systems on static images with a considerable domain gap from their synthetic counterpart. However, the validation of a CNN-based navigation system comes with additional challenges that stem from the fact that it is hard to recreate representative relative trajectories and realistic illumination conditions on-ground. Despite the recent advancements in the on-ground generation of realistic space imagery which led to the creation of the HIL images of the SPEED+ dataset (Park et al., 2021a) (Park et al., 2021), the recreation of rendezvous trajectories on-ground remains a demanding task that has not been fully explored in most of the mentioned testbeds. The recreation of realistic rendezvous trajectories on-ground has been showcased by some of the authors at ESTEC's GRALS facility (Pasqualetto Cassinis et al., 2022). In their setup, a v-bar trajectory around a mockup of the Envisat spacecraft was recreated by controlling both the target rotation and the translation of the monocular camera through two robotic arms, thus extending the testbed capability from the generation of static datasets to actual relative trajectories. However, only a static sun lamp was used to illuminate the target, leading to simplified illumination conditions. Furthermore, the calibration of the overall system was complicated by an inaccurate calibration of the target mockup, leading to lower accuracies in the reference pose labels compared to the TRON setup. In a recent effort to solve the above issues, the capabilities of the TRON facility were extended towards the generation of the Satellite Hardware-In-the-loop Rendezvous Trajectories (SHIRT) dataset (Park and D'Amico, 2022), a dataset of synthetic and realistic imagery of two close-proximity trajectories around the Tango spacecraft with high-fidelity calibration and more realistic illumination conditions of a typical LEO orbit (Park and D'Amico, 2022).

Building on the previous findings of some of the authors (Pasqualetto Cassinis et al., 2022; Park et al., 2021a; Pasqualetto Cassinis et al., 2021), this paper aims at solving the above-mentioned challenges in a sequential fashion. First, a data augmentation pipeline centered on light augmentation is introduced to solve the domain shift problem. An existing technique, proposed by Sakkos et al. (2019) for terrestrial applications, is exploited to generalize the illumination conditions during the CNN training. The performance of the CNN-based pose estimation system is then evaluated on realistic imagery of the SPEED+ dataset at a pose estimation level. Next, the CNN system is combined with an Unscented Kalman Filter (UKF) to address the performance of the proposed system at a navigation filter level. To this end, the SHIRT dataset is used in the evaluation in order to compare the filter performance in synthetic scenarios with the performance on realistic lab imagery. To cope with the challenges in the representation of the measurements uncertainty, an adaptive scheme is proposed in which the measurement error covariance is estimated online by scaling the heatmaps-based representation based on the filter innovations.

In summary, the main contributions of this work are:

1. To propose a novel data augmentation scheme based on light randomization in order to improve the CNN performance on realistic space imagery
2. To propose a new covariance adaptation method based on CNN heatmaps to capture measurements uncertainty within the navigation filter
3. To validate a tightly-coupled, CNN-based UKF on realistic trajectories generated on-ground using the TRON facility.

The paper is organized as follows. Section 2 introduces the proposed navigation framework. The TRON testbed and the image acquisition procedure are described in Section 3. In Section 4, the CNN training, validation and testing phases are detailed, with special focus to the data augmentation pipeline. Section 5 presents the preliminary pose estimation results on SPEED+ test datasets. Next, the proposed navigation filter is described in Section 6 together with the novel adaptive scheme for the measurement error covariance. Sections 7,2,3,4,5,6,7,8 describe the simulation environment and show the navigation results on the two recreated close-proximity trajectories. Finally, Section 9 provides the main conclusions and recommendations.



(a) PnP problem (Figure adapted from [3]) (b) Co-moving LVLH frame [40]

Fig. 1. Schematic of the pose estimation problem using a monocular image (a) and representation of the relative motion framework (b) (Curtis, 2005).

2. System framework

From a high-level perspective, a model-based monocular pose estimation system receives as input a 2D image and matches it with an existing wireframe 3D model of the target spacecraft to estimate the target pose with respect to the servicer camera. As illustrated in Fig. 1a, the pose estimation problem consists in determining the position of the target's centre of mass $\mathbf{t}^C$ and its attitude with respect to the camera frame C, represented by the inverse of the rotation matrix $\mathbf{R}_B^C$. The Perspective-n-Points (PnP) equations,

$$\mathbf{r}^C = [x^C \quad y^C \quad z^C]^T = \mathbf{R}_B^C \mathbf{r}^B + \mathbf{t}^C \tag{1}$$

$$\mathbf{p} = (u_i, v_i) = \left(\frac{x^C}{z^C} f_x + C_x, \frac{y^C}{z^C} f_y + C_y \right), \tag{2}$$

relate the unknown pose with a feature point $\mathbf{p}$ in the image plane via the relative position $\mathbf{r}^C$ of the feature with respect to the camera frame. Here, $\mathbf{r}^B$ is the point location in the 3D model, expressed in the body-frame coordinate system B, whereas f_x and f_y denote the focal lengths of the camera and (C_x, C_y) is the principal point of the image.

From these equations, it can already be seen that an important aspect of estimating the pose resides in the capability of the IP system to extract features $\boldsymbol{p}$ from a 2D image of the target spacecraft, which in turn need to be matched with pre-selected features $\boldsymbol{r}^B$ in the wireframe 3D model. Notably, a wireframe model of the target needs to be made available prior to the estimation. Notice also that the problem is not well defined for $n < 3$ feature points, and can have up to four positive solutions for $n = 3$ (Fischer and Bolles, 1981). Generally, more features are required in presence of large noise or symmetric objects.

The validation pipeline of the proposed navigation system is shown in Fig. 2 and consists of the following main stages:

1. **Calibration Procedure and Image Acquisition**: The adopted monocular camera is calibrated with respect to a scaled 1:2 mockup model of the Tango spacecraft via a dedicated procedure, in order to associate accurate reference labels of the pose between the camera and the mockup. Once the setup is calibrated, laboratory images of the target mockup are generated by the camera and associated to their corresponding pose labels. Trajectory design and execution is carried out in order to create a representative laboratory database (SPEED+) as well as to perform rendezvous trajectories around the mockup (SHIRT). The calibration procedure and the trajectory generation pipeline are taken from Park et al. (2021a) and Park and D'Amico (2022), respectively.
2. **Dataset Augmentation and CNN Training**: A keypoints-based CNN is trained and validated on augmented SPEED+ datasets. The augmentation is performed by introducing image noise, random background and random illuminations into the synthetically-generated images of SPEED+. The synthetic images are created using the OpenGL-based graphics renderer.
3. **Online Inference**: The keypoints-based CNN is tested on both synthetic and HIL test images. For the SPEED+ images, the pose is estimated by feeding a PnP solver with the detected keypoints as well as with the intrinsic camera parameters and the 3D model of the Tango spacecraft. Conversely, for the SHIRT images the detected keypoints are fed into a CNN-based UKF to

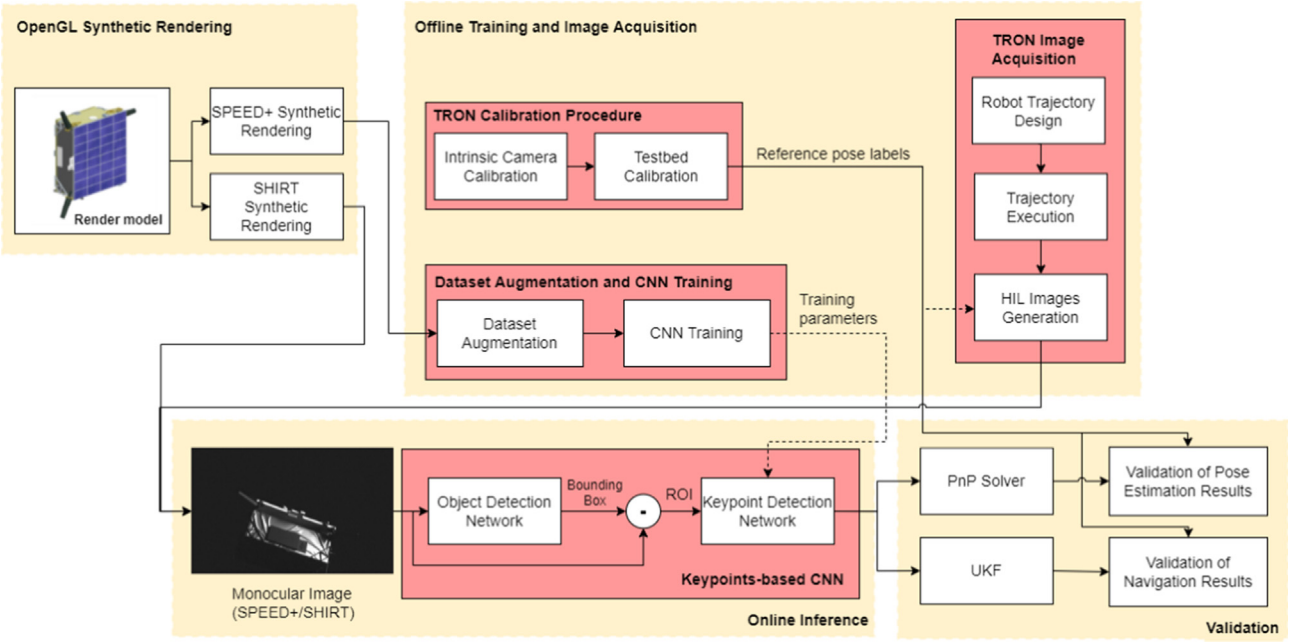


Fig. 2. Illustration of the proposed on-ground validation of the CNN-based pose estimation system.

additionally estimate the relative translational and rotational velocities.

4. **Post-Processing and Validation**: The results of the proposed CNN-based system on the HIL images are validated against the reference pose labels, derived from the calibration setup. This is performed at both pose estimation and navigation levels.

2.1. Relative navigation

This work considers a servicer spacecraft flying relative to a target spacecraft, with the relative motion being described in a Local Vertical Local Horizontal (LVLH) reference frame co-moving with the servicer (Fig. 1b). Furthermore, it is assumed that the servicer is equipped with a single monocular camera. The relative attitude of the target with respect to the servicer can then be defined as the rotation of the target body-fixed frame B with respect to the servicer camera frame C, where these frames are fixed to each spacecraft's body. The vector from the origin of the camera frame to the origin of the target frame defines their relative position. Together, these two quantities characterize the pose. This information can then be transformed from the camera frame to the servicer's center of mass by accounting for the pose of the camera with respect to the LVLH frame. Beside the Cartesian representation of the relative motion between the servicer and target spacecraft, the relative state can also be parametrized as a function of the absolute orbital elements of the two spacecraft. This paper uses the Relative Orbital Elements (ROE) state introduced by D'Amico (2010), which are defined in terms of the classical Keplerian orbital elements as:

$$\boldsymbol{\delta\alpha} = \begin{bmatrix} \delta a \\ \delta\lambda \\ \delta e_x \\ \delta e_y \\ \delta i_x \\ \delta i_y \end{bmatrix} = \begin{bmatrix} (a_t - a_s)/a_s \\ (M_t - M_s) + (\omega_t - \omega_s) + c_{i_s}(\Omega_t - \Omega_s) \\ e_t c_{\omega_t} - e_s c_{\omega_s} \\ e_t s_{\omega_t} - e_s s_{\omega_s} \\ i_t - i_s \\ s_{i_s}(\Omega_t - \Omega_s) \end{bmatrix} \tag{3}$$

where the subscripts t and s indicate the target and servicer spacecraft respectively, and s_i and c_w represent the sine and cosine of the argument of perigee ω and inclination i, respectively. Notice that this set of ROEs is nonsingular for non equatorial orbits of arbitrary eccentricity, but is singular for equatorial orbits.

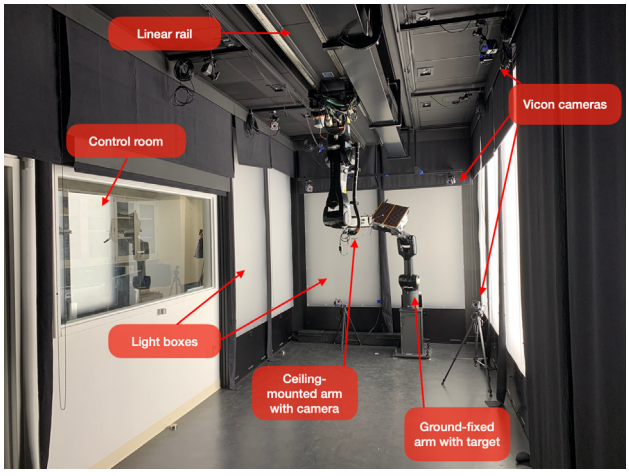


Fig. 3. TRON simulation room and its components Park et al. (2021b).

3. TRON testbed

The TRON facility at SLAB, visualized in Fig. 3, includes a control room and an 8 × 3 × 3 m simulation room which consists of various components and machineries to 1) simulate the vision-based rendezvous trajectory of a servicer spacecraft with a camera to a target spacecraft, and 2) emulate the high-fidelity spaceborne illumination conditions to maximize the realism of the images captured by the camera. TRON comprises two 6 degrees-of-freedom KUKA robot arms and a set of Vicon motion track cameras to reconfigure an arbitrary pose between a camera and a target mockup model, as well as multiple Earth albedo light boxes and a sun lamp to simulate high-fidelity spaceborne illumination conditions (Park et al., 2021b). The calibration of the facility is performed via a dedicated multi-source Robot/World Hand/Eye (RWHE) calibration procedure which fuses readings from KUKA and Vicon to accurately estimate the pose of the adopted monocular camera with respect to the target mockup. Millimeter-level position accuracy and millidegree-level orientation accuracy were obtained on a subset of close-range images of the Tango mockup. The reader is referred to Park et al. (2021b) for a detailed description.

3.1. SPEED+ dataset

The SPEED+ dataset was created in the TRON facility to be used in the ongoing international Satellite Pose Estimation Challenge, with the main objective to evaluate and compare the robustness of machine learning models trained on synthetic images (Park et al., 2021a). The dataset is built upon the existing SPEED dataset (Kisantal et al., 2020) by increasing the number of synthetic and HIL images, whilst extending the illumination conditions simulated in the facility and improving the accuracy of the pose labels. SPEED+ consists of synthetic images generated in OpenGL and realistic images generated in TRON with light boxes and sun lamp (Fig. 4). Fig. 5 illustrates sample SPEED+ images for the `synthetic`, `lightbox`, and `sunlamp` domains. Overall, the test images of SPEED+ extend the full orientation space and distance up to 10 m with realistic re-creation of Earth albedo and direct sunlight present in spaceborne imagery.

Fig. 4. Visualization of the TRON facility for the generation of the SPEED+ images. Light boxes (L1 - L10) and sunlamp locations (S1 - S3) are noted Park et al. (2021a).

3.2. SHIRT dataset

A detailed overview of the generation of the SHIRT dataset can be found in Ref. Park and D'Amico (2022). For the HIL images, the first step is to design a rendezvous trajectory which resembles a typical close-proximity scenario. Next, the motion of the two KUKA robots is commanded in order to recreate the desired trajectory, taking into account the scale of the target mockup and the constraints of the facility. In order to simulate representative illumination conditions, the *true* location of the Sun is used to intermittently switch the light boxes on/off and capture the correct inclination of the Sun with respect to the mockup. Fig. 6 shows a comparison of synthetic and HIL images for the same poses, where the synthetic images were generated for the same illumination and trajectory inputs. As can be seen, the TRON illumination manages to capture the correct inclination of the Sun used in the synthetic renderings. Besides, a clear domain gap is present between synthetic and HIL images, ensuring that the validation of the proposed CNN-based system can be carried out in challenging domain shifts. Specifically, by comparing the first (a-b) and second (c-d) rows of Fig. 6, it can be seen that domain gaps can result in either a difference in the illuminated side of the target (a-b) or in a different contrast between the illuminated and shadowed parts of the target (c-d).

4. Convolutional neural network

The main reason for an emerging interest in CNNs for feature extraction lies in the capability of the convolutional layers to extract high-level features of objects with improved robustness against image noise and illumination conditions as compared to standard IP algorithms (Pasqualetto Cassinis et al., 2021). As shown in Fig. 7, the first essential step of keypoints-based CNN systems is represented by an Object Detection Network (ODN), e.g. Faster R-CNN Ren et al. (2017), R-FCN Ren et al. (2016) or MobileNet Howard et al. (2017), placed before the main CNN. The ODN regresses the coordinates of a bounding box around the target object, in order to crop a Region Of Interest (ROI) and to increase robustness to scale, variation, and background textures. The cropped ROI is then fed into a Keypoint Detection Network, which convolves with the input image and outputs a set of feature maps. These so-called *heatmaps* are detected around pre-selected features on the target object, such as corners or interest points. The 2D pixel coordinates of the heatmap's peak intensity characterize the predicted feature location, with the intensity and the shape indicating the confidence of locating the corresponding keypoint at this position (Pavlakos et al., 2017). Notably, the selection of the

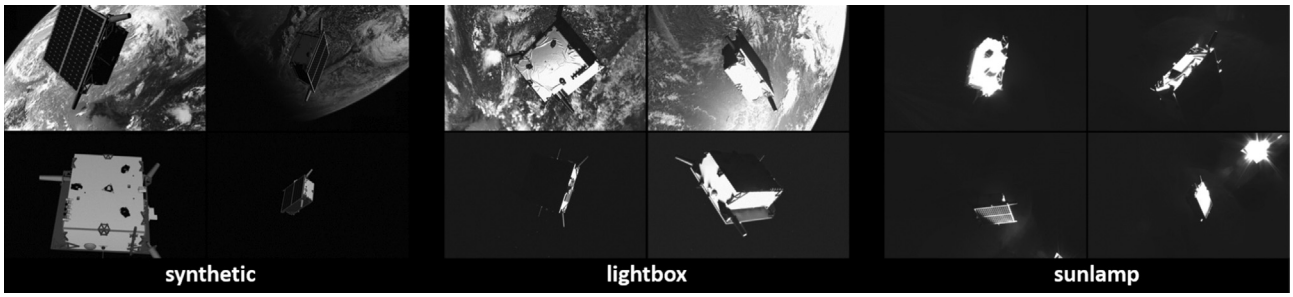


Fig. 5. Example images from different domains of SPEED+ Park et al. (2021a).

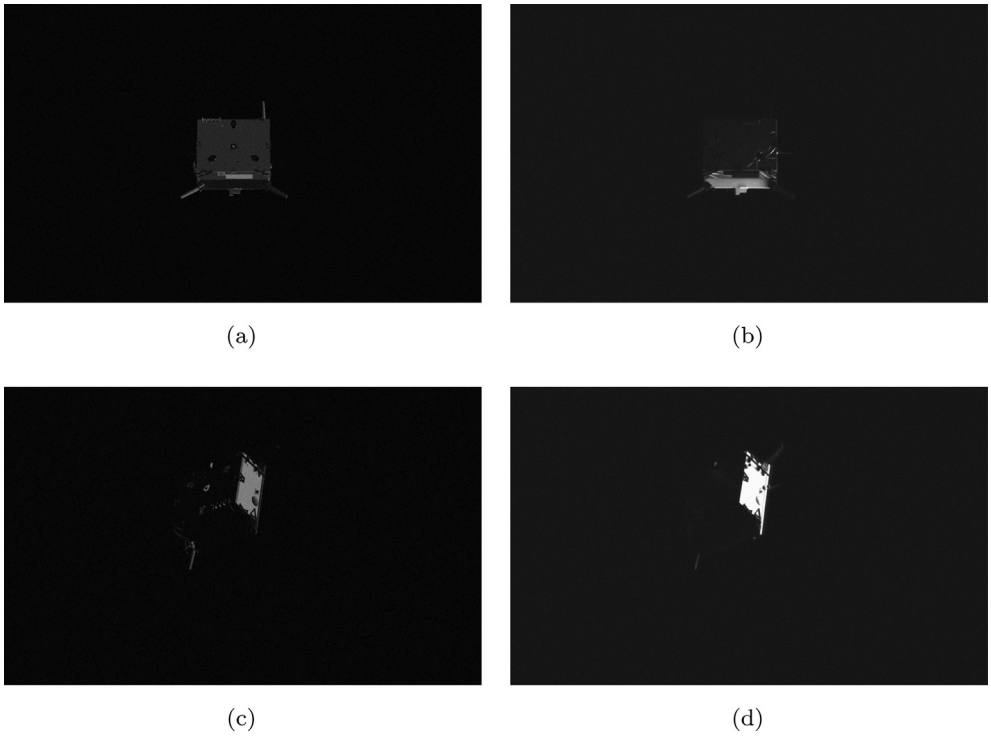


Fig. 6. Synthetic (Left) and HIL (Right) sample images for two representative poses (a-b/c-d).

CNN will drive the achievable keypoints detection accuracy and robustness. Some architectures, such as the stacked Hourglass (Newell et al., 2016) and the U-Net (Ronneberger et al., 2015), perform a downsampling of the input followed in series by an upsampling, in order to detect features at different scales. However, recent advances (Chen et al., 2019) demonstrated that by using parallel subnetworks across multiple resolutions, rather than multi-resolution serial stages, the CNN can manage to maintain a richer feature representation, facilitating more accurate and precise heatmaps. For this reason, the HRNet Sun et al. (2019) architecture is chosen in the proposed pose estimation system. The corners of the main spacecraft body and the extremities of its three antennas are pre-selected as keypoint features and used to train the CNN. These features are chosen since they already proved reliable in recent CNN trainings performed on the Tango spacecraft (Kisantal et al., 2020; Chen et al., 2019).

4.1. Data augmentation pipeline

In Fig. 8, the first step of the proposed pipeline for the datasets augmentation and randomization consists in taking the ideal synthetic images of the Tango spacecraft from the SPEED+ dataset, which already includes images with the Earth in the background. Similar to Pasqualetto Cassinis et al. (2022), a noise pipeline is then applied in order to augment the training and validation datasets with the following noise models:Fig. 9.

1. Gaussian, shot, impulse and speckle noise.
2. Gaussian, defocus, motion and zoom blurs.
3. Spatter, color jitter and random erase.

Finally, light augmentation is introduced to generalize the illumination conditions during training. The proposed approach uses the data augmentation method introduced

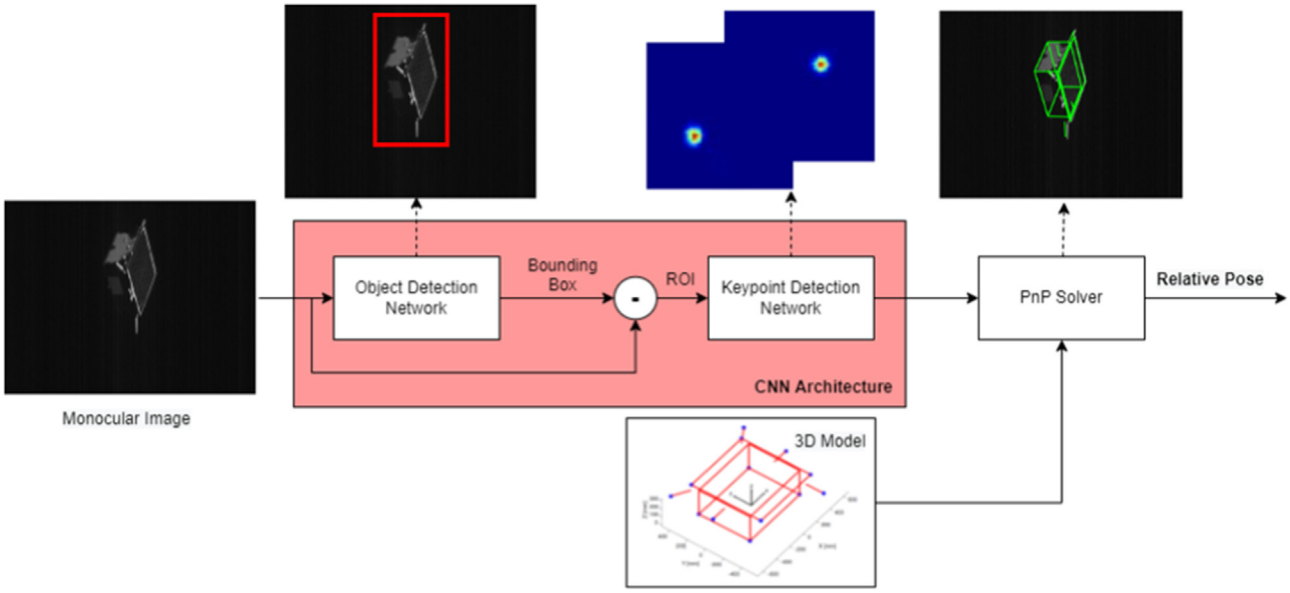


Fig. 7. Proposed CNN architecture and interface with the PnP solver.

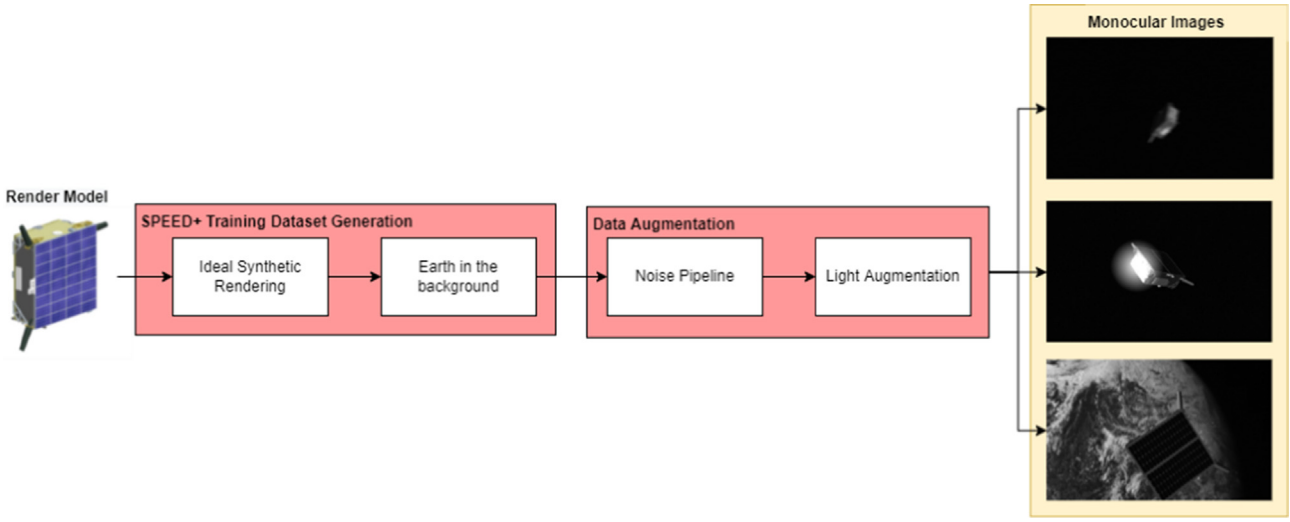


Fig. 8. SPEED+ Dataset Augmentation Pipeline.

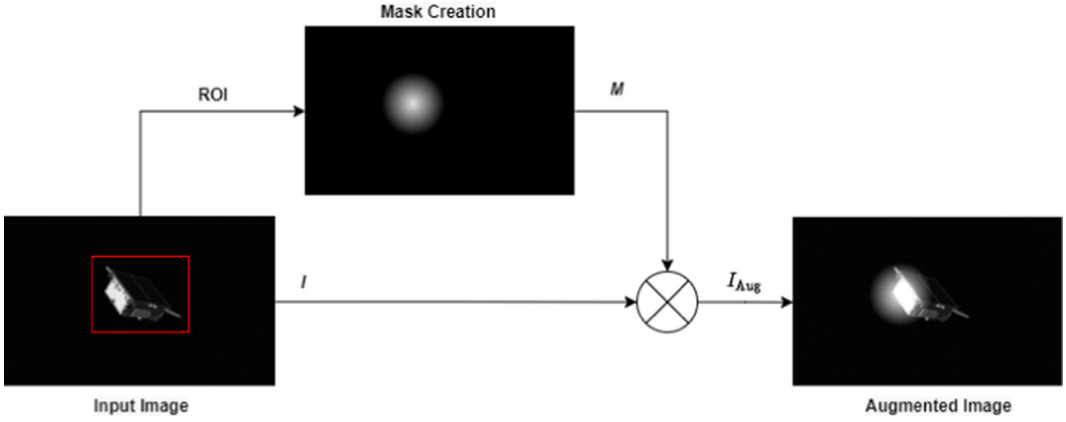


Fig. 9. Description of light augmentation.

by Sakkos et al. (2019) for terrestrial applications and extends it to the SPEED+ dataset. Fig. 8 illustrates the light augmentation pipeline for a sample image I of the SPEED+ dataset. First, a ROI is extracted around the target spacecraft and used to create an illumination mask M at random locations around the target. The mask consists of an illumination circle with centre $p = I(w,h), w \in W, h \in H$ and diameter $d = k \times \min(W,H)$, where H, W are the height and width of the extracted ROI and $k \in \left(\frac{1}{5}, \frac{1}{2}\right)$. Since uniformly modifying all pixels within the illumination circle generate unrealistic results, the Euclidean Distance Transform (EDT) is further applied to the mask in order to model light attenuation (Sakkos et al., 2019). For each pixel in the circle, the EDT assigns a number that is the distance between that pixel and the nearest nonzero pixel.

Once the mask is created, the original image is augmented to create three distinct effects:

$$\boldsymbol{I}_{\text{Aug}} = \begin{cases} \boldsymbol{I} + \boldsymbol{M}z_1 & \text{Bright filter} \\ \boldsymbol{I} - \boldsymbol{M}z_1 & \text{Dark filter} \\ \boldsymbol{I} + \boldsymbol{M}z_1 + z_2 & \text{Local and global filters} \end{cases} \tag{4}$$

where $\boldsymbol{I}_{\text{Aug}}$ is the augmented image, $\boldsymbol{M}$ is a mask of the same size as the input image $\boldsymbol{I}$, and z_1, z_2 are random integers. Fig. 10 illustrates the three filters applied to sample image I. Notably, several illumination effects can be introduced in the datasets by interchanging these filters whilst varying the disc diameter and location. In the current implementation, 50% of the synthetic images are augmented with light augmentation, of which 50% have a dark filter, 25% a bright filter, and 25% local and global filters.

4.2. Train, validation and test

Table 1 lists the number of images used in the Train, Validation, and Test datasets together with the data augmentation breakdown. The ideal synthetic images of the SPEED+ dataset are already split into 80:20 train/validation sets. The augmentation pipeline described in Section 4.1 is used to extend both sets. During training, the validation dataset is used beside the training dataset to compute the validation losses and avoid overfitting. The Adam optimizer (Kingma and Ba, 2015) is used with a cosine decaying learning rate with initial value of 10^{-3} and decaying factor of 0.1. Finally, the CNN is tested on the lightbox and sunlamp sets. In this way, the performance of a CNN trained solely on synthetic imagery can be assessed on the realistic imagery simulated in TRON.

5. Pose estimation

Following the promising pose estimation results achieved in ADR scenarios in recent studies (Kisantal et al., 2020; Chen et al., 2019; Sharma and D'Amico, 2017; Sharma and D'Amico, 2015), the Efficient Perspective-n-Points (EPnP) method followed by Gauss–Newton refinement (Lepetit and Moreno-Noguer, 2009) is selected to estimate the pose from a set of detected features. This method solves the PnP problem in Eqs. (1, 2) in closed-form with the EPnP algorithm, and uses the estimated pose as an initial guess for an iterative pose refinement. Notably, the CNN-based extraction of pre-defined features allows a direct match with the wireframe model, without the need of a large search-space for the 2D/3D correspondences which characterizes the more standard feature extractors described in Section 1.

5.1. Error metric

To evaluate the performance of the pose estimation system, two separate error metrics are generally adopted, in accordance with Kisantal et al. Kisantal et al. (2020). Firstly, the translational error E_T between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}^C$ is computed as

$$E_T = \left\| \boldsymbol{t}^C - \hat{\boldsymbol{t}}^C \right\|. \tag{5}$$

Secondly, the attitude error E_R is measured in terms of the Euler axis-angle error between the estimated quaternion $\hat{\boldsymbol{q}}$ and the ground truth $\boldsymbol{q}$,

$$\boldsymbol{\beta} = [\beta_s \quad \boldsymbol{\beta}_v] = \boldsymbol{q} \otimes \hat{\boldsymbol{q}} \tag{6}$$

$$E_R = 2\arccos(|\beta_s|), \tag{7}$$

where $\otimes$ denotes the quaternion multiplication and the subscripts s, v refer to the scalar an vectorial part of the error quaternion $\boldsymbol{\beta}$. Furthermore, a combined score is created by combining both position and attitude errors,

$$E_{\text{pose}} = E_R + \frac{E_T}{\|\boldsymbol{t}^C\|}. \tag{8}$$

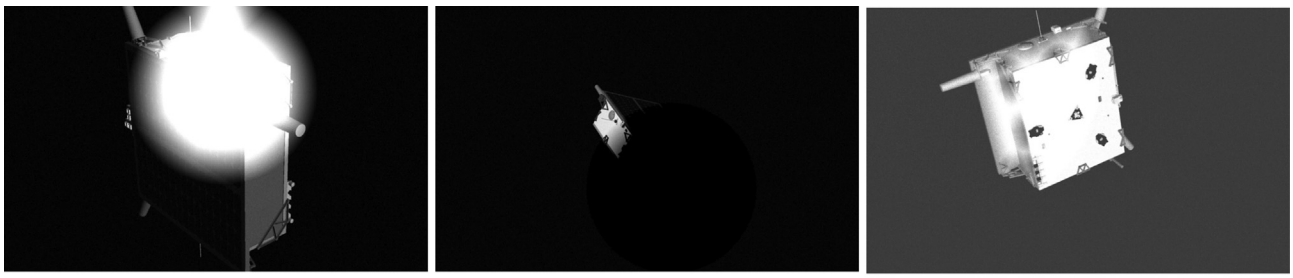


Fig. 10. Example of different light augmentation effects.

Table 1
Description of Train, Validation and Test datasets together with data augmentation breakdown.

Dataset	synthetic	lightbox	sunlamp
Train	47,966 (80% Noise Pipeline) (50% Light Augmentation)	-	-
Validation	11,994 (80% Noise Pipeline) (50% Light Augmentation)	-	-
Test	-	6,740	2,791

Note that when evaluated on SPEED+ `lightbox` and `sunlamp` samples, the modified SPEED score E^*_{pose} (Park et al., 2021a) is used to zero out the errors smaller than the thresholds based on the TRON calibration, i.e., for individual sample,

$$E^*_{\text{pose}} = \begin{cases} 0 & \text{if} \quad E_R < 0.169^\circ \quad \text{and} \quad \frac{E_T}{\|\boldsymbol{t}^C\|} < 2.173\text{mm/m} \\ E_{\text{pose}} & \text{otherwise} \end{cases} \tag{9}$$

5.2. Pose estimation results

The pose estimation results of the CNN-based system in Fig. 7 are presented for the SPEED+ test dataset, in order to evaluate the capability of the CNN to bridge the HIL imagery. Additionally, the proposed pose estimation system is evaluated on the PRISMA25 dataset, which consists of 25 flight images of the Tango spacecraft from the PRISMA mission (D'Amico et al., 2013). Despite the limited number of images in PRISMA25, the comparative study between SPEED+ and actual flight images allows an assessment of the scalability of the proposed data augmentation method to different target domains, whilst providing insight to the applicability of HIL images as a surrogate of flight images for validation.

5.2.1. SPEED+

Figs. 11, 12 show the pose estimation results in terms of the Cumulative Distribution Function (CDF) across the test datasets. Referring to Section 4.1, the results are reported for an augmentation-free CNN training, a training with the noise pipeline, and a training with both noise and light augmentation. This is done in order to assess the impact of each augmentation on the CNN performance. As can be seen, the introduction of light augmentation into the CNN training greatly improves the overall pose estimation performance on both the lightbox and the sunlamp subsets. Specifically, pose errors $E_T < 0.1\|\boldsymbol{t}^C\|, E_R < 10^\circ$ and $E_T < 0.1\|\boldsymbol{t}^C\|, E_R < 13^\circ$ are achieved in 80% of the lightbox and sunlamp subsets, respectively (Figs. 11, 12). Fig. 13 illustrates highly accurate pose estimation results on a representative subset of HIL images with lightbox illuminations, Earth in the background, and sunlamp illuminations. Conversely, Fig. 14 illustrates two representative scenarios characterized by large pose errors. As can be seen, near-eclipse illumination conditions and highly adverse sunlamp reflections can still jeopardize the CNN performance, despite the adopted data augmentation pipeline. Notably, similar effects can be observed in some of the highly accurate pose estimates (Fig. 13), suggesting that the CNN performance could be affected by small visual artifacts not visible by the human-eye.

5.2.2. PRISMA25

Fig. 15 shows the CDF pose estimation results for the PRISMA25 dataset in terms of the total SPEED score in Eqn. 9. The results are compared with the lightbox and sunlamp scenarios in order to assess the scalability of the proposed light augmentation on flight images. Notably,

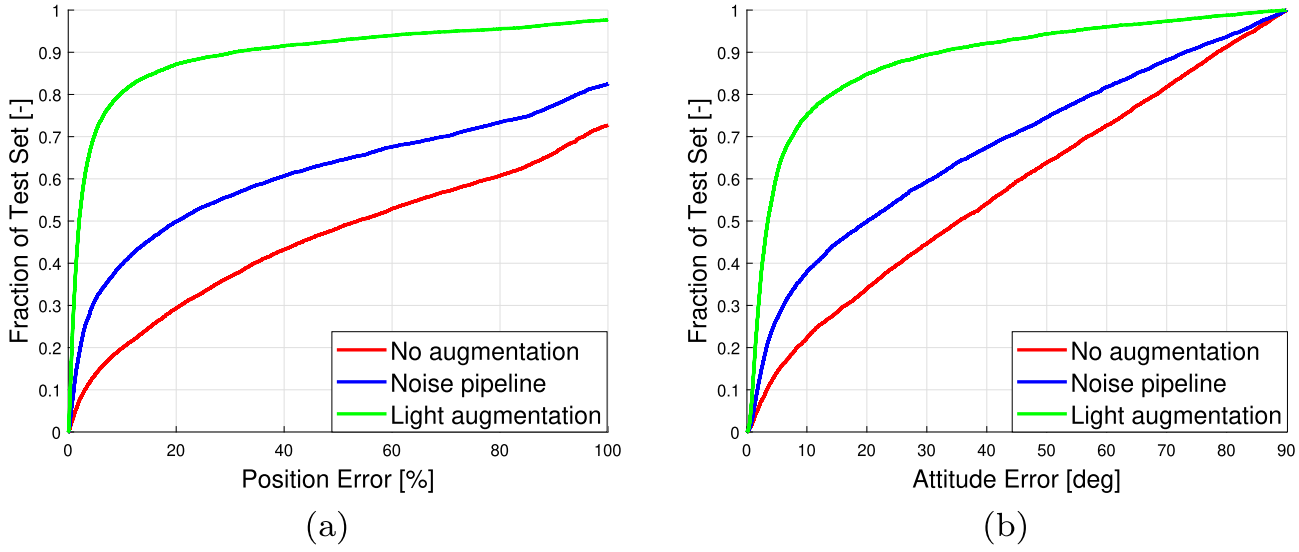


Fig. 11. Cumulative Distribution Function (CDF) of Position (a) and Attitude (b) errors over the `lightbox` subset of SPEED+.

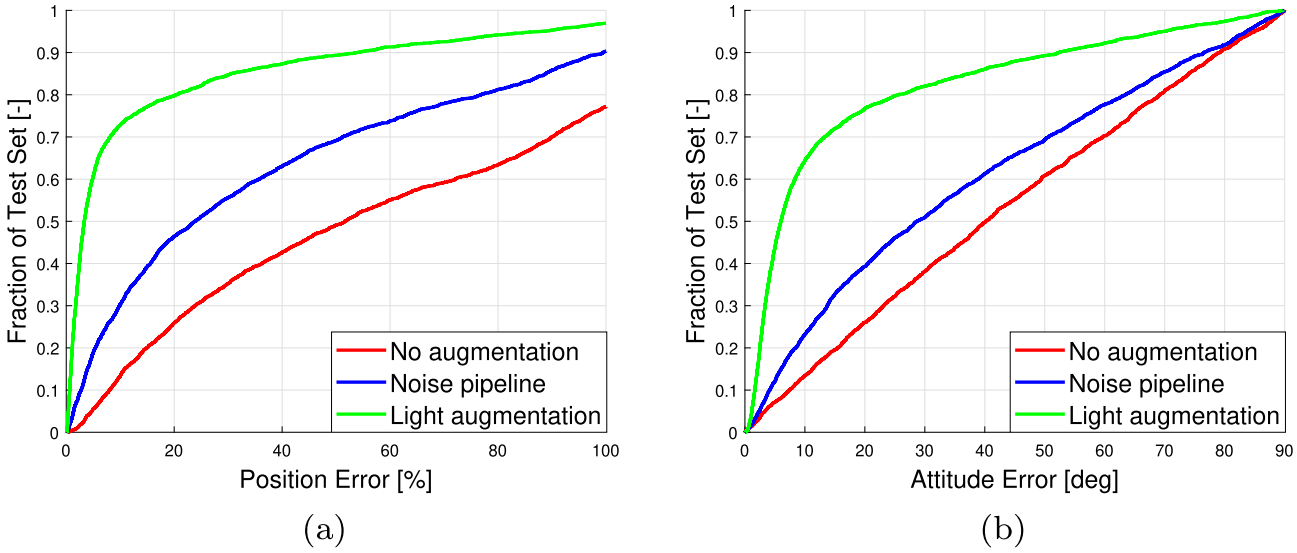


Fig. 12. Cumulative Distribution Function of Position (a) and Attitude (b) errors over the `sunlamp` subset of SPEED+.

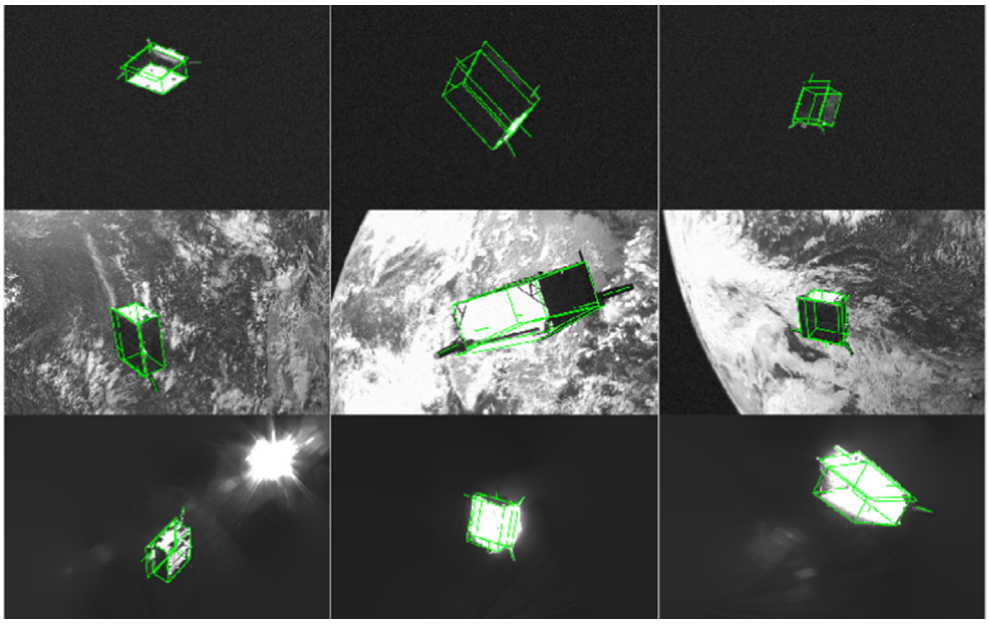


Fig. 13. Reprojection of Tango's wireframe model based on the estimated pose for subsets of the `lightbox` (Top), `lightbox` with Earth in the background (Middle) and `sunlamp` (Bottom) images.

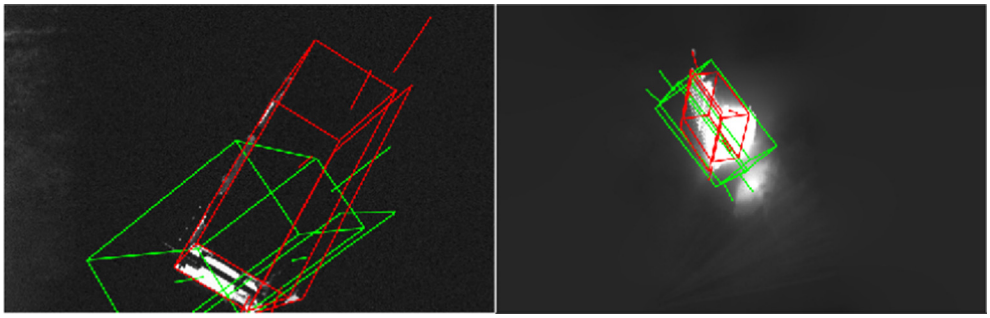


Fig. 14. Example of inaccurate pose estimates due to near-eclipse illuminations (Left) and adverse sunlamp reflections (Right). The wireframe model is projected based on the true (red) and estimated (green) poses.

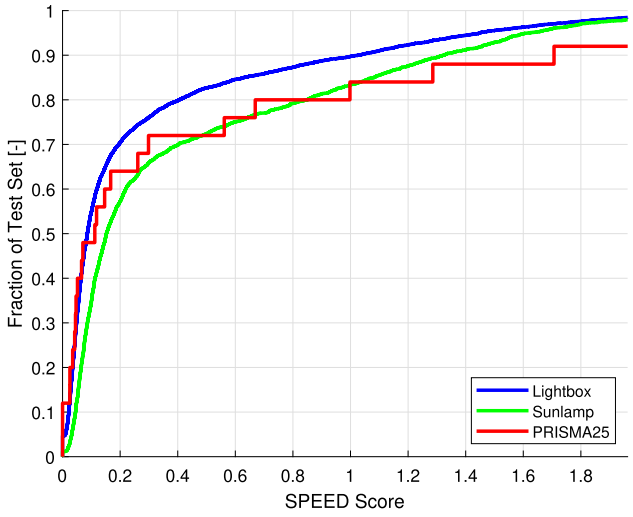


Fig. 15. E^*_{pose} in PRISMA25 compared to `lightbox` and `sunlamp`.

60% of the images are characterized by highly accurate poses with mean errors $E_T = 0.5$ m, $E_R = 2.5°$, proving the effectiveness of the proposed augmentation pipeline. Notice also that the CDF of PRISMA25 is bounded by the lightbox and sunlamp CDFs for most of the scores, suggesting that the TRON illuminations represent realistic best/worst case bounds for illuminations in actual space imagery.

Besides, Table 2 lists the mean results associated to the CDF in Fig. 15, whilst comparing the proposed light augmentation method with another augmentation pipeline which exploits texture randomization (Jackson et al., 2018). As can be seen, light augmentation provides a considerable improvement over texture randomization.

6. Navigation filter

In earlier works (Pasqualetto Cassinis et al., 2021; Sharma and D'Amico, 2017), the so-called Multiplicative Extended Kalman Filter (MEKF) was used to cope with the challenges in the attitude parametrization, which either suffers from the singularity (e.g. Euler angles) or normalization (e.g. quaternion). The latter can happen because a quaternion representation may cause the associated filter error covariance to be singular, since the quaternion set consists of four parameters to describe the 3DOF attitude. The MEKF, introduced for the first time by Lefferts et al. (1982), aims at solving the above issue by using two different parametrizations of the relative attitude. A three element error parametrization, expressed in terms of quaternions, is propagated and corrected inside the filter to return an estimate of the attitude error. At each estimation step, this error estimate is used to update a reference quaternion and is reset to zero for the next iteration. Notably, the reset step prevents the attitude error parametrization from reaching singularities, which generally occur for large angles. However, despite its promising results on ideal rendezvous scenarios with a high measurement frequency and synthetic images as measurements, the MEKF could considerably decrease its performance under low measurement frequencies and realistic space imagery, which is the case for the scenarios investigated in this paper.

In the proposed navigation system, a UKF Julier and Uhlmann (2004) is used instead, in order to capture the system nonlinearity via the unscented transform, which uses a finite set of deterministic samples instead of linearizing the nonlinearity itself. Specifically, the Unscented Quaternion Estimator (USQUE) introduced by Crassidis and Markley (2012) is adopted. The USQUE has proven more robust under low measurement acquisition frequency when compared to a more standard EKF/MEKF. Furthermore, the expected error is generally lower than the EKF, and the unscented transform avoids the derivation of Jacobian matrices (Crassidis and Markley, 2012). Similar to the MEKF, a three-component attitude-error vector is used to represent the quaternion error vector.

In a standard EKF, the state vector for pose estimation based on ROE is a 13-dimension vector composed of the relative ROE as well as the relative quaternion and rotational velocity,

$$\boldsymbol{x} = \begin{bmatrix} a_s \boldsymbol{\delta\alpha}^T & \boldsymbol{q}_B^{C\,T} & \boldsymbol{\omega}_{B/C}^{C\,T} \end{bmatrix}^T, \tag{10}$$

where a_s represents the semi-major axis of the servicer spacecraft, $\boldsymbol{q}_B^C = [q_0 \quad \boldsymbol{q}_v]$ is the quaternion set that represents the relative attitude, and $\boldsymbol{\omega}_{B/C}^C$ is the angular velocity of the target with respect to the camera, expressed in the camera frame. In the USQUE, the modified state vector propagated inside the filter becomes a 12-dimension vector,

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} a_s \boldsymbol{\delta\alpha}^T & \boldsymbol{\delta p}^T & \boldsymbol{\omega}_{B/C}^{C\,T} \end{bmatrix}^T, \tag{11}$$

where $\boldsymbol{\delta p}$ is four times the Modified Rodrigues Parameters (MRP) $\boldsymbol{\sigma}$,

$$\boldsymbol{\delta p} = 4\boldsymbol{\sigma} = 4\frac{\boldsymbol{q}_v}{1+q_0}. \tag{12}$$

Table 2
Pose estimation results on PRISMA25 dataset.

CNN Model	Data Augmentation	E^*_{pose}	E_T [m]	E_R [deg]
KRN Park et al. (2021a)	None	1.43	4.03	59.7
	Texture Rand.	1.98	7.02	77.5
HRNet	None	1.76	11.17	42.1
	Image Noise	0.96	4.22	32.1
	Image Noise/Light Rand.	**0.43**	**2.1**	**13.5**

A high-level description of the proposed navigation pipeline is provided in Fig. 16. Each subsystem is described in the following sections.

6.1. Measurement error covariance computation

Compared to the methods discussed in Section 1 (Cui et al., 2019; Harvard et al., 2020; Ferraz et al., 2014a), the proposed method derives a measurement error covariance matrix associated to each feature directly from the heatmaps detected by the CNN, rather than from the computation of the image gradient around each feature. In this method, the ith non-zero pixel around the heatmap's peak is extracted to derive a covariance matrix $\boldsymbol{C}_i$,

$$\boldsymbol{C}_i = \begin{bmatrix} \mathrm{cov}(x,x) & \mathrm{cov}(x,y) \\ \mathrm{cov}(y,x) & \mathrm{cov}(y,y) \end{bmatrix}, \tag{13}$$

where

$$\mathrm{cov}(x,y) = \sum_{i=1}^{n} w_i (x_i - p_x) \cdot (y_i - p_y). \tag{14}$$

Here, w_i is a normalized weight based on the gray intensity I_i at each pixel location, n is the number of pixels in each feature's heatmap, x, y are the coordinates of each heatmap's pixel, and $\boldsymbol{p} = (p_x, p_y)$ is the heatmap's peak location. Note that an assumption is made that the spread around the peak can represent the feature covariance, although the spread of the heat around the peak would give the expected value of the square of the map instead. Fig. 17 shows the overall flow to obtain the covariance matrix from the CNN heatmaps. The reader is referred to Pasqualetto Cassinis et al. (2021) for an exhaustive description of each step of the method.

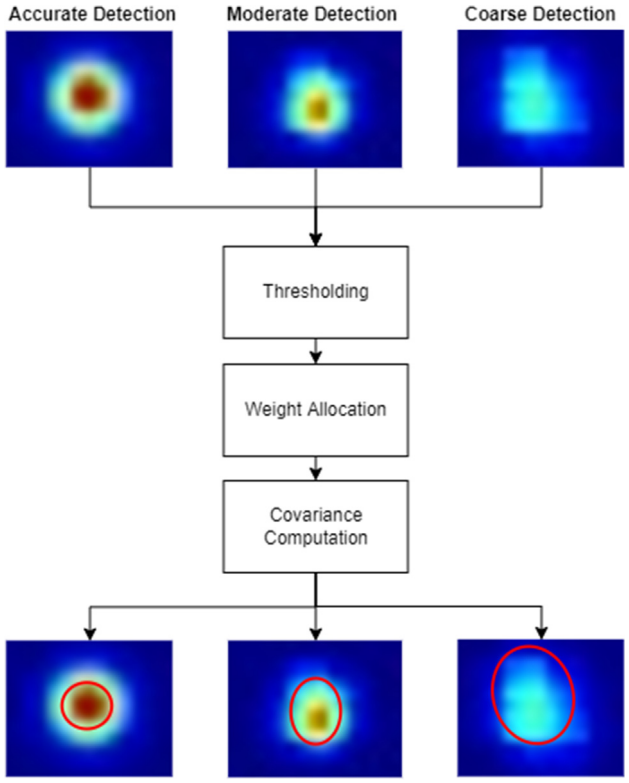


Fig. 17. Derivation of covariance matrices from CNN heatmaps. Accurate, moderate and coarse detections are represented. The displayed ellipses are derived from the computed covariances by assuming a confidence interval $\boldsymbol{1\sigma = 0.68}$.

6.2. Prediction

The first step of the filter is to generate sigma points $\boldsymbol{\chi}^{[i]}$ from the current state vector $\tilde{\boldsymbol{x}}_k$ by using the standard formulation of the UKF Julier and Uhlmann (2004):

$$\boldsymbol{\chi}_k^{[i]} = \begin{cases} \tilde{\boldsymbol{x}}_k & i = 0 \\ \tilde{\boldsymbol{x}}_k + \left(\sqrt{(N+\lambda)\boldsymbol{P}}\right)_i & i = 1, \ldots, N \\ \tilde{\boldsymbol{x}}_k - \left(\sqrt{(N+\lambda)\boldsymbol{P}}\right)_i & i = N+1, \ldots, 2N \end{cases} \tag{15}$$

where $N = 12$ is the dimensionality of the system and λ is a scaling factor tuned offline. The sigma points are then propagated through the dynamic equations. For the translational motion, the ROE state is propagated assuming an unperturbed Keplerian orbit,

$$\boldsymbol{\delta\alpha}_k = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1.5n\Delta t & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{\delta\alpha}_{k-1} \tag{16}$$

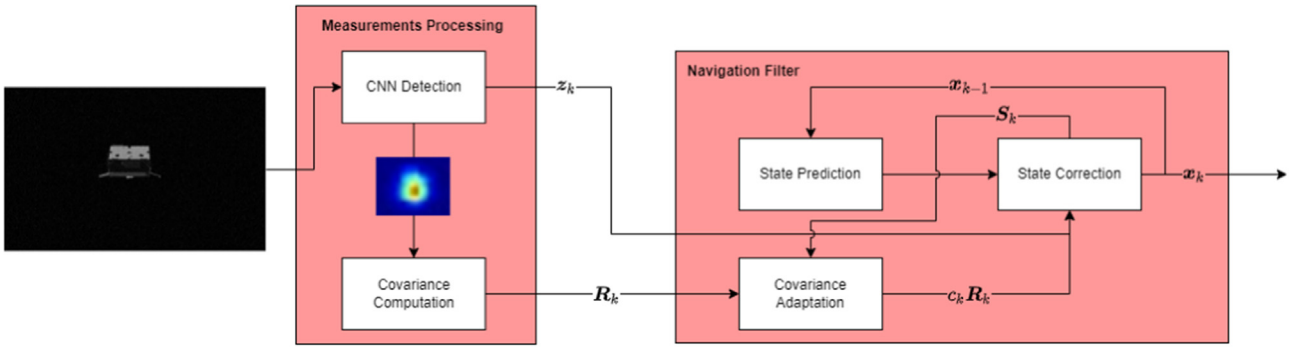


Fig. 16. High-level description of the UKF flow together with the measurements processing and adaptive covariance.

For the attitude, each sigma sample $\boldsymbol{\delta p}^{[i]}$ is transformed to the error-quaternion $\boldsymbol{\delta q}^{[i]}$, which is used to generate the quaternion multiplicatively via $\boldsymbol{q}_k^{+,[i]} = \boldsymbol{\delta q}^{[i]} \otimes \boldsymbol{q}_k$, where $\otimes$ denotes the quaternion multiplication. The dynamics update of the quaternion representation of relative orientation assumes perturbation-free motion,

$$\boldsymbol{q}_k^- = \begin{bmatrix} \cos\left(\frac{1}{2}\left\|\boldsymbol{\omega}_{B/C}^C\right\|\Delta t\right) \\ \hat{\boldsymbol{\omega}}\sin\left(\frac{1}{2}\left\|\boldsymbol{\omega}_{B/C}^C\right\|\Delta t\right) \end{bmatrix} \boldsymbol{q}_{k-1}^+ \tag{17}$$

where Δt is the time-step between filter calls and $\hat{\boldsymbol{\omega}}$ is the normalized angular velocity, representing the direction of the target rotation axis relative to the servicer. Afterwards, $q_k^{-,[i]}$ is used to compute $dp_k^{-,[i]}$. The mean of dp_k^- provided by the unscented transform is used to update the nominal state quaternion.

The relative angular velocity $\boldsymbol{\omega}_{B/C}^C$ is propagated via the following equation (Capuano et al., 2020):

$$\begin{aligned}\dot{\boldsymbol{\omega}}_{B/C}^C = \boldsymbol{R}_B^C\left(\boldsymbol{\tau}_B - \boldsymbol{\omega}_{B/E}^B \times \boldsymbol{J}_B\boldsymbol{\omega}_{B/E}^B\right) \\ - \boldsymbol{J}_C^{-1}\left(\boldsymbol{\tau}_C - \boldsymbol{\omega}_{C/E}^C \times \boldsymbol{J}_C\boldsymbol{\omega}_{C/E}^C\right) - \boldsymbol{\omega}_{C/E}^C \times \boldsymbol{\omega}_{B/C}^C\end{aligned} \tag{18}$$

where $\boldsymbol{J}$ is the target's inertia matrix, $\boldsymbol{\tau}$ are the external torques, and E denotes the Earth-Centered-Inertial (ECI) frame. Note that in the current implementation no control torques are modeled for neither the target nor the servicer, i.e. $\boldsymbol{\tau}_B = \boldsymbol{\tau}_C = \boldsymbol{0}$. After propagation, the sigma points are used to derive a mean state estimate and its associated error covariance,

$$\bar{\boldsymbol{x}}_k = \sum_{i=0}^{2N} w^{[i]}\boldsymbol{\chi}^{[i]} \tag{19}$$

$$\boldsymbol{P}_k = \sum_{i=0}^{2N} w^{[i]}\left(\boldsymbol{\chi}^{[i]} - \bar{\boldsymbol{x}}_k\right)\left(\boldsymbol{\chi}^{[i]} - \bar{\boldsymbol{x}}_k\right)^T + \boldsymbol{Q}_k \tag{20}$$

where the weights $w^{[i]}$ are a function of λ, N and $\boldsymbol{Q}_k$ is the process noise covariance.

6.3. Correction

The measurement update follows the projections described in Eqs. (1, 2), in which the sigma points of the relative position $\boldsymbol{t}^C$ are derived from the ROE state given the knowledge of the servicer's orbital elements state and its attitude with respect to the ECI frame. Once the expected measurements $\boldsymbol{Z}^{[i]}$ given each sigma point are computed, the mean expected measurement and innovation covariance can be computed similar to the mean and covariance of the filter state,

$$\bar{\boldsymbol{z}}_k = \sum_{i=0}^{2N} w^{[i]}\boldsymbol{Z}^{[i]} \tag{21}$$

$$\boldsymbol{S}_k = \sum_{i=0}^{2N} w^{[i]}\left(\boldsymbol{Z}^{[i]} - \bar{\boldsymbol{z}}\right)\left(\boldsymbol{Z}^{[i]} - \bar{\boldsymbol{z}}\right)^T + \boldsymbol{R}_k \tag{22}$$

where $\boldsymbol{R}_k$ represents the measurement error covariance matrix. In the proposed system, $\boldsymbol{R}_k$ is a time-varying block diagonal matrix constructed with the heatmaps-derived covariances $\boldsymbol{C}_i$ in Eq. (13),

$$\boldsymbol{R}_k = c_k\begin{bmatrix} \boldsymbol{C}_1 & & \\ & \ddots & \\ & & \boldsymbol{C}_n \end{bmatrix}. \tag{23}$$

where the scaling coefficient c_k is described in Section 6.4. Notice that $\boldsymbol{C}_i$ can differ for each feature in a given frame as well as vary over time. Preliminary navigation results (Pasqualetto Cassinis et al., 2021; Pasqualetto Cassinis et al., 2020) already showed that such heatmaps-derived covariance matrix can capture the statistical distribution of the measured features and improve the measurement update step of the navigation filter.

At this stage, outliers are removed from the mean measurements by computing the Mahalanobis Distance M_i between the ith feature and its corresponding filter innovation term $\boldsymbol{\Delta}_k^z$,

$$M_i = \sqrt{\boldsymbol{\Delta}_{i,k}^z \boldsymbol{S}_k^{-1} \boldsymbol{\Delta}_{i,k}^{z\,T}} \tag{24}$$

where $\boldsymbol{\Delta}_k^z = \bar{\boldsymbol{z}} - \boldsymbol{h}(\bar{\boldsymbol{x}}_k)$ and $\boldsymbol{h}$ is the nonlinear transformation in Eqs. (1, 2). The Mahalanobis distance is a measure of the distance between a point and a distribution. In this case, the point is the keypoint detected by the CNN in the image and the distribution is the reprojected feature and its associated covariance. The threshold M_t to select outliers is determined by

$$M_t = \sqrt{-2\ln(p_m)} \tag{25}$$

where p_m is the desired probability that a measurement would result in $M_i \geqslant M_t$, given that the correspondence between the detected and reprojected features is correct (Thrun et al., 2005). In other words, if $M_i \geqslant M_t$ it is highly unlikely that the ith keypoint correlates to its reprojection as predicted by the filter, and the feature can be rejected. In this way, filter robustness can be improved during low visibility periods of the target spacecraft in which wrong CNN detections may occur. Notably, this feature rejection scheme reflects an important advantage of incorporating a navigation filter compared to relying solely on the CNN detection and PnP solver solution.

Finally, the corrected state estimate $\hat{\boldsymbol{x}}_k$ is obtained from the propagated state $\tilde{\boldsymbol{x}}_k$, the innovation $\boldsymbol{\Delta}_k^z$, and the Kalman Gain $\boldsymbol{K}_k$,

$$\hat{\boldsymbol{x}}_k = \bar{\boldsymbol{x}}_k + \boldsymbol{K}_k\boldsymbol{\Delta}_k^z. \tag{26}$$

where $\boldsymbol{K}_k$ is a function of the state error and innovation covariance, and $\boldsymbol{K}_k\boldsymbol{\Delta}_k^z$ represents the state correction.

After correction, the new attitude error $\boldsymbol{\delta p}$ is reset to zero at each iteration.

6.4. Covariance adaptation

The heatmaps-based covariance described in Section 6.1 already proved to accurately represent the measurements uncertainty of the CNN detections in synthetic images with ideal illumination conditions (Pasqualetto Cassinis et al., 2021). However, a reliable representativeness of the heatmaps cannot always be guaranteed in realistic space images under challenging illumination conditions. Despite capturing the shape of the distribution, the magnitude of the heatmpas-derived covariance could indeed fail at representing the actual detection uncertainty of the CNN. This could lead the navigation filter to trust inaccurate features and ultimately diverge.

Referring to Fig. 16, the above challenges are addressed by adaptively estimating the measurement error covariance through a new technique which leverages the heatmaps-based covariance together with existing covariance matching techniques. This approach is based on the adaptive state noise compensation method for estimating the process noise covariance introduced in Stacey and D'Amico (2021).

By replacing the theoretical covariance of the innovations, also known as pre-fit residuals, in a Kalman filter with an empirical estimate, the measurement error covariance of the ith pair of pixel measurements at time step k can be estimated as Myers (1974)

$$\hat{\mathbf{R}}_{i,k} = \boldsymbol{\Delta}^{z}_{i,k}\boldsymbol{\Delta}^{z^T}_{i,k} - \bar{\mathbf{S}}_{i,k}. \tag{27}$$

Here, the innovation $\boldsymbol{\Delta}^{z}_{i,k}$ is the difference between the true and expected ith pair of pixel measurement at time step k, taking into account all measurements through time step $k-1$. The matrix $\bar{\mathbf{S}}_k = \mathbf{S}_k - \mathbf{R}_k$ is computed from the sigma points passed through the nonlinear measurement models in Eq. (22) (Thrun et al., 2005). The portion of $\bar{\mathbf{S}}_k$ corresponding to the ith pair of pixel measurements is $\bar{\mathbf{S}}_{i,k} \in \mathbb{R}^{2\times 2}$. Eq. (27) assumes a filter at steady state and that the pixel measurement errors are zero-mean. Typically, Eq. (27) is averaged over some finite length sliding window of filter output. Such covariance matching techniques, also referred to as innovation based estimation, are not guaranteed to converge to the true estimate of the error covariance. However, they are widely used and have been shown to work well in practice (Myers and Tapley, 1976; Sullivan and D'Amico, 2017; Karlgaard, 2010; Fraser and Ulrich, 2021; Mohamed and Schwarz, 1999).

Assuming that the pixel measurement errors of different features are not correlated with each other and that the associated heatmaps are representative of the shape of their covariance yields the model

$$\mathbf{R}_{i,k} = c_k\mathbf{C}_{i,k}. \tag{28}$$

Here, c_k is introduced to account for the uncertainty in the magnitude of the covariance. Using all the pixel measurements at a single filter call, a pseudo-coefficient c^*_k can be estimated through a weighted least squares fit between the diagonal elements of the right hand side of Eq. (28) and the corresponding elements of Eq. (27). The resulting solution for c^*_k is

$$c^*_k = \underset{c}{\operatorname{argmin}}\ ||\mathbf{W}^{-1/2}_k(\boldsymbol{A}_k c - \boldsymbol{b}_k)||^2 \tag{29}$$

$$= \frac{\boldsymbol{A}^T_k\mathbf{W}^{-1}_k\boldsymbol{b}_k}{\boldsymbol{A}^T_k\mathbf{W}^{-1}_k\boldsymbol{A}_k}. \tag{30}$$

The vector $\boldsymbol{b}_k$ is the concatenation of all the main diagonal elements of the covariance matching estimates $\hat{\mathbf{R}}_{i,k} \in \mathbb{R}^{2\times 2}$ for all i at time step k. The vector $\mathbf{A}_k$ contains the corresponding diagonal elements of each $\mathbf{C}_{i,k}$.

The weighting matrix $\mathbf{W}_k$ is chosen as the theoretical covariance of $\boldsymbol{b}_k$ such that covariance matching estimates with less uncertainty have a greater influence on the solution of c^*_k. The covariance between any two elements of $\boldsymbol{b}_k$ is

$$\operatorname{Cov}\left(\widehat{R}^{\alpha}_{i,k}, \widehat{R}^{\beta}_{j,k}\right) = \operatorname{Cov}\left(\Delta^{\alpha^2}_{i,k}, \Delta^{\beta^2}_{j,k}\right) \tag{31}$$

$$= \mathrm{E}\left[\Delta^{\alpha^2}_{i,k}\Delta^{\beta^2}_{j,k}\right] - \mathrm{E}\left[\Delta^{\alpha^2}_{i,k}\right]\mathrm{E}\left[\Delta^{\beta^2}_{j,k}\right] \tag{32}$$

$$= 2\mathrm{E}\left[\Delta^{\alpha}_{i,k}\Delta^{\beta}_{j,k}\right]^2. \tag{33}$$

The scalar $\widehat{R}^{\alpha}_{i,k}$ is either the first or second element of the main diagonal of $\hat{\mathbf{R}}_{i,k}$ as denoted by α. Eq. (33) is just an element of $\mathbf{S}_k$ squared multiplied by two. Thus,

$$\mathbf{W}_k = \mathbf{S}^{\circ 2}_k \tag{34}$$

where the factor of 2 from Eq. (33) has been dropped because it does not change the result of Eq. (30). The Hadamard power, $^{\circ}$, denotes an element-wise power. Eq. (34) assumes that $\boldsymbol{b}_k$ is ordered in the same way as the pixel measurements are ordered in the measurement vector provided to the filter. For computational efficiency, the weighting matrix is approximated as diagonal by setting all off-diagonal elements to zero. As a result, the inverse of the weighting matrix required in Eq. (30) can be computed element-wise along the main diagonal.

The variance of a weighted least squares estimate (Gill and Montenbruck, 2012) is

$$P_{c^*_k} = \operatorname{Var}\left(c^*_k\right) \tag{35}$$

$$= \left(\boldsymbol{A}^T_k\mathbf{W}^{-1}_k\boldsymbol{A}_k\right)^{-1}. \tag{36}$$

The final estimate of c_k is obtained by combining the c^* estimates over a sliding window of length N through

$$c_k = \underset{c}{\operatorname{argmin}}\ ||\bar{\mathbf{W}}^{-1/2}_k\left(\bar{\boldsymbol{A}}_k c - \bar{\boldsymbol{b}}_k\right)||^2 \tag{37}$$

$$= \frac{\bar{\boldsymbol{A}}^T_k\bar{\mathbf{W}}^{-1}_k\bar{\boldsymbol{b}}_k}{\bar{\boldsymbol{A}}^T_k\bar{\mathbf{W}}^{-1}_k\bar{\boldsymbol{A}}_k} \tag{38}$$

where

$$\bar{\boldsymbol{A}}_k = \boldsymbol{1}_{N\times 1}, \quad \bar{\boldsymbol{b}}_k = \begin{bmatrix} c_k^* & \dots & c_{k-N+1}^* \end{bmatrix}^T \tag{39}$$

$$\bar{\mathbf{W}}_k = \begin{bmatrix} P_{c_k^*} & 0 & \dots & 0 \\ 0 & P_{c_{k-1}^*} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & P_{c_{k-N+1}^*} \end{bmatrix}. \tag{40}$$

Upper and lower inequality constraints are easily added to Eq. (37) by setting the estimated c_k equal to any constraint that is violated. Note that some lower bound greater than zero should always be included to guarantee that the measurement error covariance is positive definite. Any filter calls where no pixel measurements are provided to the filter are excluded from the sliding window. The weighting matrix in Eq. (40) is assumed diagonal because the innovations are not correlated in time for a filter at steady state (Mehra, 1972; Kailath, 1968). After each measurement update, $c_k^*, P_{c_k^*}$, and c_k are computed through Eqs. (30), (36), and (38) respectively. The resulting c_k is used in the following measurement update to compute the measurement error covariance of each pixel measurement through Eq. (28).

7. Simulations

The investigated rendezvous scenarios of the SHIRT dataset involve the Mango and Tango spacecraft from the PRISMA mission (D'Amico et al., 2013). The simulations take place in Low Earth Orbit (LEO) with initial ROEs listed in Table 3. ROE 1 is purely an along-track separation describing a standard v-bar hold point, whereas ROE 2 describes a mid-range hold point (Sharma and D'Amico, 2017), meaning that the relative trajectory slightly drifts away from the v-bar. The reference truth motion of each spacecraft is numerically propagated with 1 s interval using rigorous force models including GRACE's GGM05S geopotential of order and degree 120 (Ries et al., 2016), NRLMSISE-00 atmospheric model Picone et al. (2002), analytic lunisolar third-body gravity (Gill and Montenbruck, 2012), and solar radiation pressure (SRP) including a conical Earth shadow model. Moreover, the spacecraft attitude is perturbed via analytical gravity-gradient (Wertz, 1978), air drag, and SRP effects. The satellite parameters for these forces match the modeled parameters of the Mango and Tango spacecraft specified in D'Amico (2010). The servicer's angular velocity is initialized to always align the camera boresight with the along-track direction, whereas the target's initial angular velocity about its principal axes is set to $\boldsymbol{\omega}_0 = [1 \quad 0 \quad 0]^T(^\circ/s)$ for ROE1 and $\boldsymbol{\omega}_0 = [0 \quad 0.4 \quad -0.6]^T(^\circ/s)$ for ROE2. The images are captured every 5 s. However, a measurement acquisition time of 30 s is used in this work, in order to represent a severely limited on-board processing power. Fig. 18 shows a montage of the HIL images for both ROE trajectories.

8. Results

In this section, the navigation results are presented for the ROE 1 and ROE 2 trajectories. Similar to the pose estimation error metrics introduced in Section 5.1, the norms of the estimated translational and rotational velocities are compared to their ground truth values to return the estimation error. Table 4 reports the obtained navigation results for both synthetic and HIL scenarios in terms of mean error at steady state. The mean values after two orbits are computed across a time interval of 600 s at steady state.

8.1. Synthetic scenarios

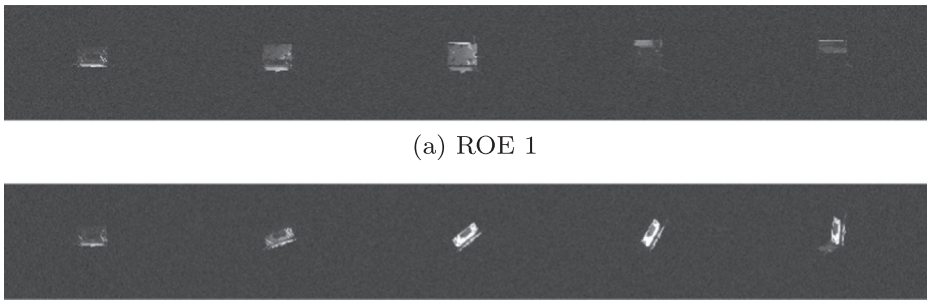
Figs. 19, 20 show the navigation performance on both relative trajectories when synthetic images are used as measurements. Results are reported for the pose estimation error along with the standard deviation derived from the state covariance matrix. In these scenarios, centimeter-level position error and degree-level attitude error are achieved at steady state after two relative orbits. Also, notice that the state covariance increases at approximately half of each relative orbit before quickly decreasing. In these periods, some of the CNN detections become inaccurate due to the challenging near-eclipse conditions in the target's visibility. However, thanks to the adaptive heatmaps-based measurement error covariance the filter can capture the detection uncertainty and reflect it in a larger state covariance. This is a desirable behaviour as it prevents the filter from diverging due to an inaccurate representation of the measurements uncertainty.

8.2. HIL scenarios

As mentioned in Section 6.4, the magnitude of the heatmaps-based covariance is not guaranteed to reflect the actual measurements uncertainty in realistic imagery, especially under highly challenging illumination conditions. As such, the state covariance could be updated with an

Table 3
Initial mean servicer orbital elements and reference relative trajectories parametrized in ROE space Park and D'Amico (2022).

Servicer Orbit	$a = 7078.1$ [km]	$e = 0.001$	$i = 98.2^\circ$	$\Omega = 189.9^\circ$	$\omega = 0^\circ$	$M_0 = 0^\circ$
Initial ROE	$a\delta a$ [m]	$a\delta\lambda$ [m]	$a\delta e_x$ [m]	$a\delta e_y$ [m]	$a\delta i_x$ [m]	$a\delta i_y$ [m]
ROE 1	0	−8	0	0	0	0
ROE 2	−0.25	−8	0	0.15	0	−0.15



(a) ROE 1

(b) ROE 2

Fig. 18. First five HIL images for ROE 1 (a) and ROE 2 (b) trajectories. The images were generated every 30 s to represent limited-on board processing power.

Table 4
Mean navigation results after two orbits for ROE 1 and ROE 2 trajectories, computed across a time interval of 600 s at steady state. Results are reported for both synthetic and HIL scenarios.

Trajectory	Measurements	E_T [m]	E_R [m]	E_v [m/s]	E_ω [°/s]
ROE 1	`synthetic`	0.13 ± 0.02	0.73 ± 0.37	1.2E-11 ± 2E-12	0.008 ± 0.005
	`lightbox`	0.25 ± 0.06	14 ± 11	1.3E-11 ± 1.6E-12	0.5 ± 0.4
ROE 2	`synthetic`	0.02 ± 0.02	0.97 ± 0.85	8.5E-12 ± 1E-12	0.01 ± 0.005
	`lightbox`	0.38 ± 0.02	9 ± 7.5	4.9E-11 ± 6.8E-12	0.3 ± 0.2

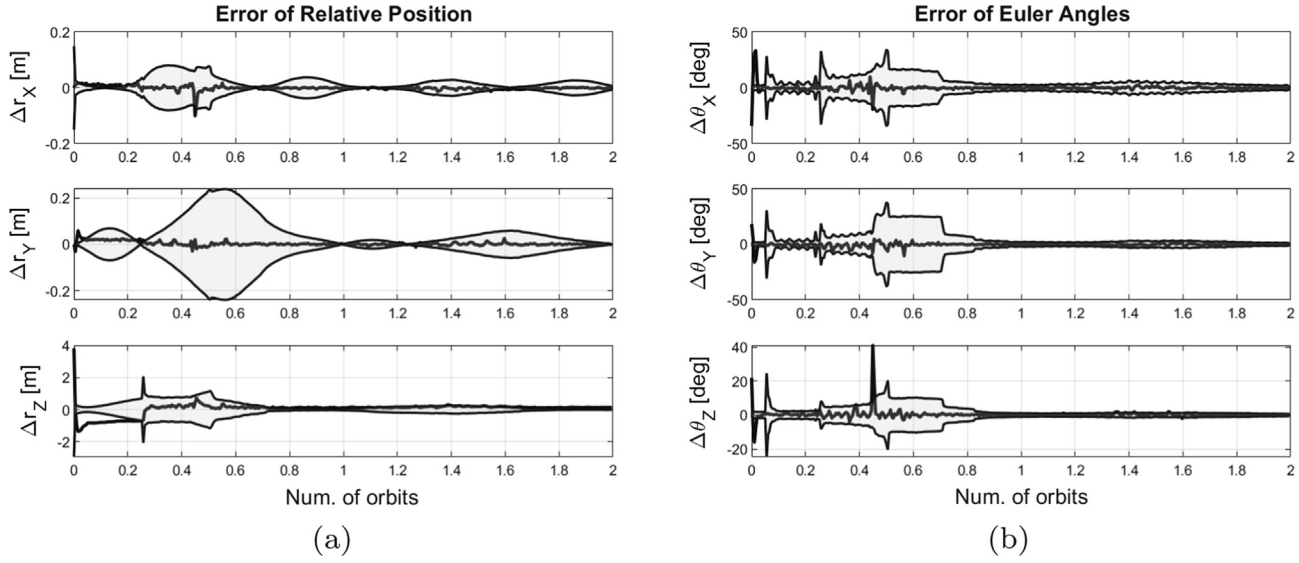


Fig. 19. Pose Estimation Error in the ROE 1 synthetic scenario. The ROE state is converted to position (left) and the quaternion state is converted to 3–2-1 Euler angles (right). The errors are shown along with the 3σ error bound. Note that the error in the Z direction is characterized and by an initial offset in the position and a sudden jump in attitude. The position offset is a result of a relatively large error in the initial a priori knowledge of the camera boresight direction. For the attitude, this could instead be a consequence of filter contamination with wrong measurements, which is not captured by the measurement covariance matrix.

inaccurate measurement error covariance (Eq. (22)), leading to the divergence of the filter immediately after initialization. This was observed for both ROE trajectories when a simple heatmaps-based covariance was used. However, the achieved performance of the filter can benefit from the proposed covariance adaptation scheme. After the estimated scaling factor is introduced (Eq. (28)), the filter shows an increase in robustness towards inaccurate measurements and does not diverge. Figs. 21, 22 show the navigation performance on both relative trajectories when the HIL images are used. Thanks to the covariance adaptation, centimeter-level position error is achieved after two relative orbits, which compares well with the results in the synthetic scenario. However, a larger uncertainty in the attitude estimate can be seen in both trajectories, with steady-state mean errors $E_T \approx 10^\circ - 15^\circ$. A degraded performance in the relative attitude due to challenging HIL images was already observed by some of the authors in a similar

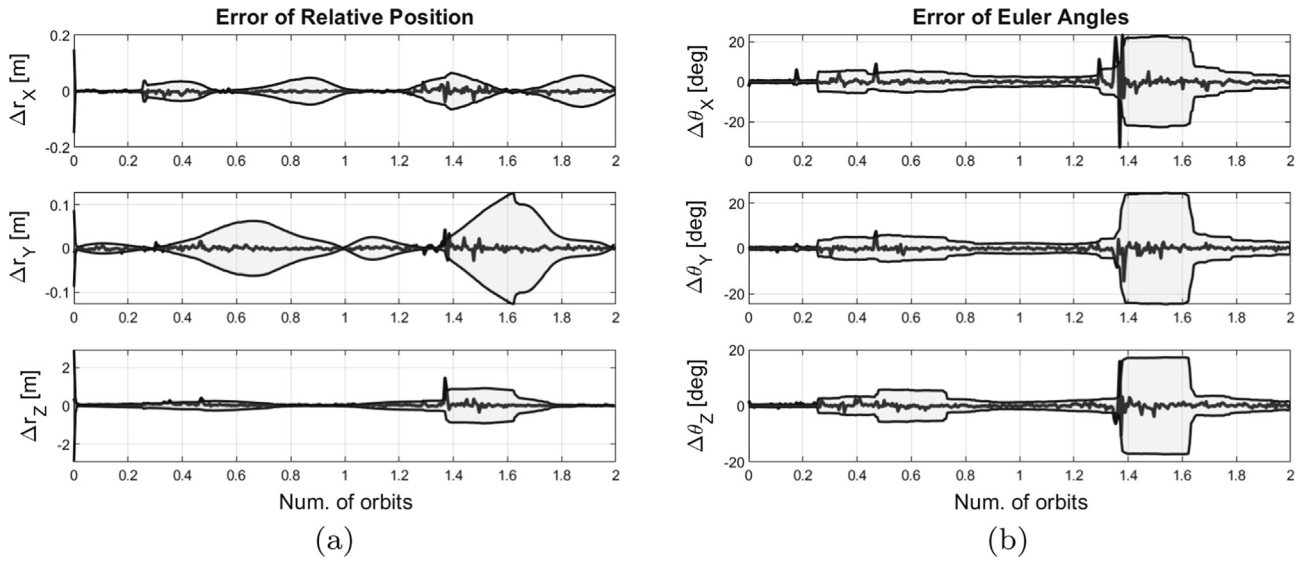


Fig. 20. Pose Estimation Error in the ROE 2 synthetic scenario. The ROE state is converted to position (left) and quaternion state to 3–2-1 Euler angles (right). The errors are shown along with the 3σ error bound.

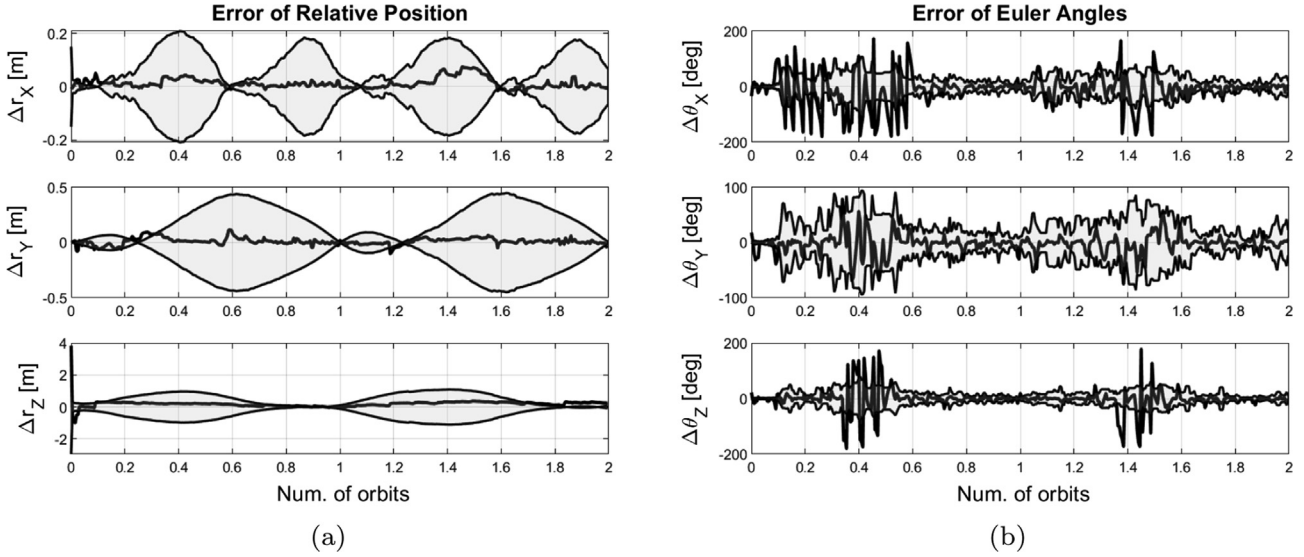


Fig. 21. Pose Estimation Error in the ROE 1 HIL scenario.

validation setup of a CNN-based pose estimation system (Pasqualetto Cassinis et al., 2022), suggesting that the domain shift affects the estimate of the rotational state more than the translational one. Nevertheless, the velocity estimates listed in Table 4 indicate that the proposed system can estimate the relative rotational velocity with relatively small errors. Note also that, due to a non-ideal tuning of the process noise covariance Q, the navigation error exceeds the 3-σ bounds in some portions of the orbits. The same behaviour is observed when varying the initial conditions for both the state and the covariance. For a more optimal tuning of the navigation filter, which goes beyond the scope of this paper, the reader is referred to Park and D'Amico (2022), in which the process noise covariance is adaptively estimated based on the filter innovation.

Overall, the gained filter robustness highlights the benefits of adaptively correcting the measurement error covariance. Notably, the proposed covariance adaptation scheme was applied to a more standard, gradient-based covariance method (Cui et al., 2019) which does not exploit the CNN heatmaps, in order to compare its performance with the heatmaps-based method. Preliminary results showed a much worse navigation performance with mean attitude errors $E_R > 100°$ after one relative orbit, suggesting that the image gradient does not correlate well with the CNN uncertainties. This indicates that not only the adaptive covariance is essential for filter robustness, but also that

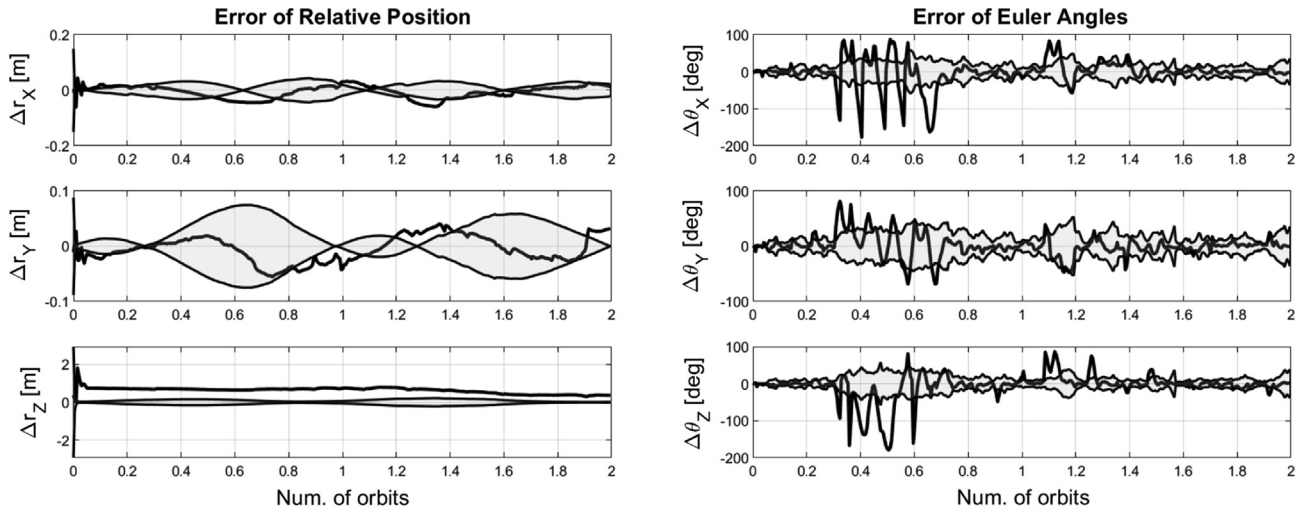


Fig. 22. Pose Estimation Error in the ROE 2 HIL scenario.

a proper representation of the measurements uncertainty is required beforehand.

9. Conclusions

This paper introduces the validation of an adaptive Convolutional Neural Network (CNN)-based navigation filter for the pose estimation of an uncooperative spacecraft by an active servicer spacecraft equipped with a monocular camera. First, the performance of the proposed system is evaluated at a pose estimation level by testing the adopted CNN on the realistic space imagery of the SPEED+ dataset, captured from Stanford's Testbed for Rendezvous and Optical Navigation (TRON). By adopting a novel data augmentation pipeline centered on light augmentation, the system is proven capable of bridging the domain gap between the synthetic training images and the HIL test images, showcasing the benefits of generalizing the illumination conditions during training. Specifically, pose errors $E_T < 0.1\|\boldsymbol{t}^C\|, E_R < 10^\circ$ and $E_T < 0.1\|\boldsymbol{t}^C\|, E_R < 13^\circ$ are achieved in 80% of the lightbox and sunlamp subsets, respectively. Furthermore, by assessing the pose estimation performance on both Hardware-In-the-Loop (HIL) images and actual space imagery from the PRISMA mission, the data augmentation method is shown to be robust against changes in the target domain.

At a navigation filter level, the system is validated by assessing the performance of the proposed CNN-based Unscented Kalman Filter (UKF) on both synthetic and HIL images of two representative rendezvous trajectories around the target spacecraft. Results on the synthetic scenarios indicate that the system can accurately estimate the relative state with centimeter- and sub-degree-level pose errors, thanks to the heatmaps-based representation of the CNN detection uncertainty. In the HIL scenarios, the inclusion of an adaptation scheme for the measurement error covariance returns centimeter-level position errors and moderate attitude accuracies ($E_R \approx 10^\circ - 15^\circ$) at steady-state, preventing the filter from diverging during periods of low measurement accuracy. Remarkably, results suggest that a proper representation of the measurements uncertainty combined with an adaptive measurement error covariance is key in improving the navigation robustness.

Recommendations for future work are identified at different levels of the proposed navigation system. First of all, other data augmentation techniques shall be investigated together with other CNN architectures, with the aim to further improve the CNN detection accuracy prior to the navigation filter. As an example, the impact of combining texture and light augmentation on the CNN performance could be explored. At a filter level, the modeling of the process noise covariance should be improved in order to better capture the uncertainty in the state estimate. Finally, the applicability of the proposed CNN-based system shall be demonstrated on a representative space processor in order to carry out a real-time performance analysis. In this context, the interface of the proposed system with guidance and control should also be investigated in closed-loop tests.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is funded and supported by the European Space Agency and Airbus Defence and Space under Network/Partnering Initiative (NPI) program with Grant No. NPI 577-2017. The Visiting Student Research program at Stanford University is funded by the Fulbright-Schuman

Program, issued by the Commission for Educational Exchange between the United States, Belgium and Luxembourg.

References

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. 110, 346–359.

Ben-David, S., Blitzer, J., Crammer, K., Pereira, F., 2007. Analysis of representations for domain adaptation. In: Scholkopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Information Processing Systems. MIT Press, Oxford.

Capuano, V., Kim, K., Harvard, A., Chung, S., 2020. Monocular-based pose determination of uncooperative space objects. Acta Astronaut. 166, 493–506.

Chen, B., Cao, J., Parra, A., Chin, T., 2019. Satellite pose estimation with deep landmark regression and nonlinear pose refinement. In: International Conference on Computer Vision, Seoul, South Korea.

Crassidis, J., Markley, F., 2012. Unscented filtering for spacecraft attitude estimation. J. Guidance Control Dyn. 26, 536–542.

Cui, J., Min, C., Bai, X., Cui, J., 2019. An improved pose estimation method based on projection vector with noise error uncertainty. IEEE Photonics J. 11.

Curtis, H., 2005. Orbital Mechanics for Engineering Students. Elsevier.

D'Amico, S., 2010. Autonomous Formation Flying in low Earth Orbit, Ph.D. thesis. Delft University of Technology.

D'Amico, S. et al., 2013. Prisma. In: Distributed Space Missions for Earth System Monitoring, pp. 599–637. https://doi.org/10.1007/978-1-4614-4541-8.

D'Amico, S., Benn, M., Jorgensen, J., 2014. Pose estimation of an uncooperative spacecraft from actual space imagery. Int. J. Space Sci. Eng. 2, 171–189.

Donahue, J., Krahenbuhl, P., Darrell, T., 2017. Adversarial feature learning. In: Int. Conf. Learning Representation, Toulon, France.

Dubanchet, V., Bejar Romero, J., Gregertsen, K., Austad, H., Gancet, J., Natusiewicz, K., Vinals, J., Guerra, G., Rekleitis, G., Paraskevas, I., Nanos, K., Papadopoulos, E., Majewski, L., Ferraris, S., Purnell, J., Casu, D., D'Amico, J., Andiappane, S., 2020. EROSS project – European autonomous robotic vehicle for on-orbit servicing. In: i-SAIRAS Virtual Conference, Virtual.

Ferraz, L., Binefa, X., Moreno-Noguer, F., 2014a. Leveraging Feature Uncertainty in the PnP Problem. In: Proceedings of the British Machine Vision Conference, Nottingham, UK. https://doi.org/10.5244/C.28.83.

Ferraz, L., Binefa, X., Moreno-Noguer, F., 2014b. Very Fast Solution to the PnP Problem with Algebraic Outlier Rejection. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. https://doi.org/10.1109/CVPR.2014.71.

Fischer, M.A., Bolles, R., 1981. Random Sample Consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24, 381–395.

Fraser, C.T., Ulrich, S., 2021. Adaptive extended Kalman filtering strategies for spacecraft formation relative navigation. Acta Astronaut. 178.

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., W.F.A., Brendel, W., 2019. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. In: International Conference on Learning Representations, New Orleans, LA, USA.

Ghifary, M., Kleijn, W., Zhang, M., Balduzzi, D., 2016. Deep reconstruction-classification networks for unsupervised domain adaptation. In: Int. Conf. Computer Vision, Rome, Italy.

Gill, E., Montenbruck, O., 2012. Satellite Orbits - Models, Methods and Applications. Springer.

Harvard, A., Capuano, V., Shao, E., Chung, S.-J., 2020. Spacecraft Pose Estimation from Monocular Images Using Neural Network Based Keypoints and Visibility Maps. In: AIAA Scitech 2020 Forum, Orlando, FL, USA. https://doi.org/10.1109/AERO.2018.8396425.

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. In: ArXiv Preprint, doi:arXiv:1704.04861.

Huo, Y., Li, Z., Zhang, F., 2020. Fast and accurate spacecraft pose estimation from single shot space imagery using box reliability and keypoints existence judgments. IEEE Access 8.

Jackson, P., A.-A.A., Bonner, S., Breckon, T., Obara, B., 2018. Style augmentation: Data augmentation via style randomization. In: Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA.

Julier, S., Uhlmann, J., 2004. Unscented filtering and nonlinear estimation. Proc. IEEE 92, 401–422.

Kailath, T., 1968. An innovations approach to least-squares estimation–Part I: Linear filtering in additive white noise. IEEE Trans. Autom. Control 13.

Karlgaard, C., 2010. Robust Adaptive Estimation for Autonomous Rendezvous in Elliptical Orbit, Ph.D. thesis. Virginia Polytechnic Institute and State University.

Kingma, D., Ba, J., 2015. Adam: A method for stochastic optimization. In: 3rd International Conference for Learning Representations, San Diego, CA, USA. https://doi.org/10.2514/6.2018-2100.

Kisantal, M., Sharma, S., Park, T.H., Izzo, D., Martens, M., D'Amico, S., 2020. Satellite Pose Estimation Challenge: Dataset, competition design and results. IEEE Trans. Aerosp. Electron. Syst. 56 (5),, 4083–4098.

Krüger, H., Theil, S., 2010. TRON - hardware-in-the-loop test facility for lunar descent and landing optical navigation. In: IFAC-ACA Automatic Control in Aerospace, Nara, Japan.

Lefferts, E., Markley, F., Shuster, M., 1982. Kalman filtering for spacecraft attitude estimation. J. Guidance Control Dyn. 5, 417–429.

Lepetit, F., Moreno-Noguer, P., 2009. Fua, EPnP: an accurate O(n) solution to the PnP problem. Int. J. Comput. Vision 81, 155–166.

Lowe, D., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision 60, 91–110.

Mehra, R., 1972. Approaches to adaptive filtering. IEEE Trans. Autom. Control 17, 693–698.

Mohamed, A.H., Schwarz, K.P., 1999. Adaptive Kalman filtering for INS/GPS. J. Geodesy 73, 193–203.

Myers, K.A., 1974. Filtering Theory Methods and Applications to the Orbit Determination Problem for Near-Earth Satellites, Ph.D. thesis. The University of Texas at Austin.

Myers, K., Tapley, B., 1976. Adaptive sequential estimation with unknown noise statistics. IEEE Trans. Autom. Control 21, 520–523.

Newell, A., Yang, K., Deng, J., 2016. Stacked hourglass networks for human pose estimation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision - ECCV 2016, vol. 9912. Springer, Cham, pp. 483–499.

Ostrowsky, A., 1966. Solution of Equations and Systems of Equations, 2nd ed. Academic Press, New York.

Park, T.H., D'Amico, S., 2022. Adaptive neural network-based unscented Kalman filter for spacecraft pose tracking at rendezvous. In: AAS/AIAA Astrodynamics Specialist Conference, Charlotte, NC, USA.

Park, T.H., Sharma, S., D'Amico, S., 2019. Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft. In: AAS/AIAA Astrodynamics Specialist Conference, Portland, ME, USA. https://doi.org/10.48550/arXiv.1909.00392.

Park, T.H., Martens, M., Jawaid, M., Wang, Z., Chen, B., Chin, T.-J., Izzo, D., D'Amico, S., 2023. Satellite Pose Estimation Competition 2021: Results and Analyses. Acta Astronaut 204, 640–665 https://doi.org/10.1016/j.actaastro.2023.01.002.

Park, T., Martens, M., Lecuyer, G., Izzo, D., D'Amico, S., 2021a. Speed +: Next-generation dataset for spacecraft pose estimation across domain gap. In: ArXiv Preprint. https://doi.org/10.48550/arXiv.2110.03101.

Park, T., Bosse, J., D'Amico, S., 2021b. Robotic testbed for rendezvous and optical navigation: Multi-source calibration and machine learning use cases. In: AAS/AIAA Astrodynamics Specialist Conference, Big Sky, MT, USA.

Pasqualetto Cassinis, L. et al., 2022. On-ground validation of a CNN-based monocular pose estimation system for uncooperative spacecraft: Bridging domain shift in rendezvous scenarios. Acta Astronaut. 196, 123–138.

Pasqualetto Cassinis, L., Fonod, R., Gill, E., 2019. Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft. Prog. Aerosp. Sci. 110.

Pasqualetto Cassinis, L., Fonod, R., Gill, E., Ahrns, I., Gil Fernandez, J., 2020. CNN-based pose estimation system for close-proximity operations around uncooperative spacecraft. In: AIAA Scitech 2019 Forum, Orlando, FL, USA. https://doi.org/10.2514/6.2020-1457.

Pasqualetto Cassinis, L., Fonod, R., Gill, E., Ahrns, I., Gil-Fernandez, J., 2021. Evaluation of tightly- and loosely-coupled approaches in CNN-based pose estimation systems for uncooperative spacecraft. Acta Astronaut. 182, 189–202.

Pavlakos, G., Zhou, X., Chan, A., Derpanis, K., Daniilidis, K., 2017. 6-DoF object pose from semantic keypoints. In: IEEE International Conference on Robotics and Automation, Marina Bay Sands, Singapore.

Picone, J., Hedin, A., Drob, D., Aikin, A., 2002. NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. J. Geophys. Res.: Space Phys. 107, SIA 15-1–SIA 15-16.

Ren, S., He, K., Girshick, R., Sun, J., 2016. Object detection via region-based fully convolutional networks. Adv. Neural Informat. Process. Syst., 379–387

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39, 1137–1149.

Ries, J., Bettadpur, S., Eanes, R.J., Kang, Z., Ko, U.-D., McCullough, C., Nagel, P., Pie, N., Poole, S., Richter, T., Save, H., Tapley, B., 2016. The development and evaluation of the global gravity model GGM05, Technical Report CSR-16-02, Center for Space Research, The University of Texas at Austin.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to sift or surf. In: International Conference on Computer Vision, Barcelona, Spain, pp. 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544.

Sakkos, D., Shum, H., Ho, E., 2019. Illumination-based data augmentation for robust background subtraction. In: Proceedings of the 2019 International Conference on Software Knowledge Information Management and Applications (SKIMA).

Sharma, S., D'Amico, S., 2015. Comparative Assessment of techniques for initial pose estimation using monocular vision. Acta Astronaut. 123, 435–445.

Sharma, S., D'Amico, S., 2017. Reduced-dynamics pose estimation for non-cooperative spacecraft rendezvous using monocular vision. In: 38th AAS Guidance and Control Conference, Breckenridge, CO, USA.

Sharma, S., D'Amico, S., 2020. Neural network-based pose estimation for noncooperative spacecraft rendezvous. IEEE Trans. Aerosp. Electron. Syst. 56.

Sharma, S., Beierle, C., D'Amico, S., 2018. Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks. In: IEEE Aerospace Conference, Big Sky, MT, USA. https://doi.org/10.1109/AERO.2018.8396425.

Sharma, S., Ventura, J., D'Amico, S., 2018. Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous. J. Spacecraft Rock. 55, 1–16.

Shi, J., Ulrich, S., Ruel, S., Simulation, CubeSat, 2018. Detection using monocular camera images and convolutional neural networks. In: 2018 AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, USA. https://doi.org/10.2514/6.2018-1604.

Sonawani, S., Alimo, R., Detry, R., Jeong, D., Hess, A., Ben Amor, H., 2020. Assistive relative pose estimation for on-orbit assembly using convolutional neural networks. In: AIAA Scitech 2020 Forum, Orlando, FL, USA. https://doi.org/10.1109/AERO.2018.8396425.

Stacey, N., D'Amico, S., 2021. Adaptive and dynamically constrained process noise estimation for orbit determination. IEEE Trans. Aerosp. Electron. Syst.

Sugiyama, M., Muller, K.-R., 2005. Input-dependent estimation of generalization error under covariate shift. Stat. Decis. 23, 249–279.

Sullivan, J., D'Amico, S., 2017. Nonlinear Kalman filtering for improved angles-only navigation using relative orbital elements. J. Guidance Control Dyn. 40, 2183–2200.

Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep high-resolution representation learning for human pose estimation. In: 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA.

Tatsch, A., Fitz-Coy, N., Gladun, S., 2006. On-orbit servicing: A brief survey. In: Proceedings of the 2006 Performance Metrics for Intelligent Systems Workshop, pp. 21–23.

Thrun, S., Burgard, W., Fox, D., 2005. Probabilistic Robotics. The MIT Press, pp. 65–71.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., 2017. Domain randomization for transferring deep neural networks from simulation to the real world. Int. Conf. Intelligent Robots and Systems, 23–30. https://doi.org/10.1109/IROS.2017.8202133.

Wertz, J., 1978. Spacecraft Attitude Determination and Control. Reidel Dordrecht.

Wieser, M., Richard, H., Hausmann, G., Meyer, J.-C., Jaekel, S., Lavagna, M., Biesbroek, R., 2015. e.deorbit mission: OHB debris removal concepts. In: ASTRA 2015–13th Symposium on Advanced Space Technologies in Robotics and Automation, Noordwijk, The Netherlands.

Wilde, M., Clark, C., Romano, M., 2019. Historical survey of kinematic and dynamic spacecraft simulators for laboratory experimentation of on-orbit proximity maneuvers. Prog. Aerosp. Sci. 110.

Zwick, M., Huertas, I., Gerdes, L., Ortega, G., 2018. ORGL - ESA's test facility for approach and contact operations in orbital and planetary environments. In: International Symposium on Artificial Intelligence, Robotics and Automation in Space, Madrid, Spain.

Sharma, S., Park, T.H., D'Amico, S., 2019. Spacecraft Pose Estimation Dataset (SPEED). In: Stanford Digital Repository. Available at https://purl.stanford.edu/dz692fn7184.https://doi.org/10.25740/dz692fn7184.

Park, T.H., Martens, M., D'Amico, S., 2021. Next Generation Spacecraft Pose Estimation Dataset (SPEED+). In: Stanford Digital Repository. Available at https://purl.stanford.edu/wv398fc4383.https://doi.org/10.25740/wv398fc4383.

Park, T.H., D'Amico, S., 2022. SHIRT: Satellite Hardware-In-the-Loop Rendezvous Trajectories Dataset. In: Stanford Digital Repository. Available at https://purl.stanford.edu/zq716br5462.https://doi.org/10.25740/zq716br5462.