Evaluation of tightly- and loosely-coupled approaches in CNN-based pose estimation systems for uncooperative spacecraft

Pasqualetto Cassinis, Lorenzo; Fonod, Robert; Gill, Eberhard; Ahrns, Ingo; Gil-Fernández, Jesús

DOI
10.1016/j.actaastro.2021.01.035

Publication date
2021

Document Version
Final published version

Published in
Acta Astronautica

Acta Astronautica 182 (2021) 189–202

Contents lists available at ScienceDirect

Acta Astronautica

journal homepage: www.elsevier.com/locate/actaastro

Research paper

Evaluation of tightly- and loosely-coupled approaches in CNN-based pose estimation systems for uncooperative spacecraft

Lorenzo Pasqualetto Cassinis [a,*], Robert Fonod [a], Eberhard Gill [a], Ingo Ahrns [b], Jesús Gil-Fernández [c]

[a] Delft University of Technology, Kluyverweg 1 2629 HS, Delft, The Netherlands
[b] Airbus DS GmbH, Airbusallee 1, 28199, Bremen, Germany
[c] ESTEC, Keplerlaan 1, 2201 AZ, Noordwijk, The Netherlands

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Relative pose estimation
Active Debris Removal
In-Orbit Servicing
Monocular-based navigation filters
Convolutional Neural Networks

ABSTRACT

The relative pose estimation of an inactive spacecraft by an active servicer spacecraft is a critical task in the design of current and planned space missions, due to its relevance for close-proximity operations, such as In-Orbit Servicing and Active Debris Removal. This paper introduces a novel framework to enable robust monocular pose estimation for close-proximity operations around an uncooperative spacecraft, which combines a Convolutional Neural Network (CNN) for feature detection with a Covariant Efficient Procrustes Perspective-n-Points (CEPPnP) solver and a Multiplicative Extended Kalman Filter (MEKF). The performance of the proposed method is evaluated at different levels of the pose estimation system. A Single-stack Hourglass CNN is proposed for the feature detection step in order to decrease the computational load of the Image Processing (IP), and its accuracy is compared to the standard, more complex High-Resolution Net (HRNet). Subsequently, heatmaps-derived covariance matrices are included in the CEPPnP solver to assess the pose estimation accuracy prior to the navigation filter. This is done in order to support the performance evaluation of the proposed tightly-coupled approach against a loosely-coupled approach, in which the detected features are converted into pseudomeasurements of the relative pose prior to the filter. The performance results of the proposed system indicate that a tightly-coupled approach can guarantee an advantageous coupling between the rotational and translational states within the filter, whilst reflecting a representative measurements covariance. This suggest a promising scheme to cope with the challenging demand for robust navigation in close-proximity scenarios. Synthetic 2D images of the European Space Agency's Envisat spacecraft are used to generate datasets for training, validation and testing of the CNN. Likewise, the images are used to recreate a representative close-proximity scenario for the validation of the proposed filter.

1. Introduction

Nowadays, key Earth-based applications such as remote sensing, navigation, and telecommunication, rely on satellite technology on a daily basis. To ensure a high reliability of these services, the safety and operations of satellites in orbit has to be guaranteed. In this context, advancements in the field of Guidance, Navigation, and Control (GNC) were made in the past years to cope with the challenges involved in In-Orbit Servicing (IOS) and Active Debris Removal (ADR) missions [1,2]. For such scenarios, the estimation of the relative pose (position and attitude) of an uncooperative spacecraft by an active servicer spacecraft represents a critical task. Compared to cooperative close-proximity missions, the pose estimation problem is indeed complicated by the fact that the target satellite is not functional and/or not able to aid the relative navigation. Hence, optical sensors shall be preferred over Radio Frequency (RF) sensors to cope with a lack of navigation devices such as Global Positioning System (GPS) sensors and/or antennas onboard the target.

From a high-level perspective, optical sensors can be divided into active and passive devices, depending on whether they require power to function, i.e. LIght Detection And Ranging (LIDAR) sensors and Time-Of-Flight (TOF) cameras, or if they passively acquire radiation, i.e. monocular and stereo cameras. Spacecraft relative navigation usually exploits Electro-Optical (EO) sensors such as stereo cameras [3,4] and/or a LIDAR sensor [5] in combination with one or more monocular cameras, in order to overcome the partial observability that results from

* Corresponding author.
E-mail addresses: L.PasqualettoCassinis@tudelft.nl (L. Pasqualetto Cassinis), Robert.Fonod@ieee.org (R. Fonod), E.K.A.Gill@tudelft.nl (E. Gill), ingo.ahrns@airbus.com (I. Ahrns), Jesus.Gil.Fernandez@esa.int (J. Gil-Fernández).

https://doi.org/10.1016/j.actaastro.2021.01.035
Received 13 June 2020; Received in revised form 2 December 2020; Accepted 21 January 2021
Available online 15 February 2021

the lack of range information in these latter [6]. In this framework, 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 [7]. However, given the low Signal-To-Noise Ratio (SNR) and the high contrast which characterize space images, a significant effort is still required to comply with most of the demanding requirements for a robust and accurate monocular-based navigation system. Interested readers are referred to Pasqualetto Cassinis et al. [8] for a recent overview of the current trends in monocular-based pose estimation systems. Notably, the aforementioned navigation system cannot rely on known visual markers, as they are typically not installed 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/or edges on the target body. 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 relative pose. In other words, a reliable detection of key features under adverse orbital conditions is highly desirable to guarantee safe operations around an uncooperative spacecraft. Moreover, it would be beneficial from a different standpoint to obtain a model of feature detection uncertainties. This would provide the navigation system with additional statistical information about the measurements, which could in turn improve the robustness of the entire estimation process.

Unfortunately, standard pose estimation solvers such as the Efficient Perspective-n-Point (EPnP) [9], the Efficient Procrustes Perspective-n-Point (EPPnP) [10], or the multi-dimensional Newton Raphson Method (NRM) [11] do not have the capability to include features uncertainties. Only recently, the Maximum-Likelihood PnP (MLPnP) [12] and the Covariant EPPnP (CEPPnP) [13] solvers were introduced to exploit statistical information by including feature covariances in the pose estimation. Ferraz et al. [13] proposed a method for computing the covariance which takes different camera poses to create a fictitious distribution around each detected keypoint. Other authors proposed an improved pose estimation method based on projection vector, in which the covariance is associated to the image gradient magnitude and direction at each feature location [14], or a method in which covariance information is derived for each feature based on feature's visibility and robustness against illumination changes [15]. 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. Moreover, this procedure could not be easily applied if Convolutional Neural Networks (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, another procedure should be followed in which the output of the CNNs is directly exploited to return relevant statistical information about the detection step. This could, in turn, provide a reliable representation of the detection uncertainty.

The implementation of CNNs for monocular pose estimation in space has already become an attractive solution in recent years, also thanks to the creation of the Spacecraft PosE Estimation Dataset (SPEED) [16], a database of highly representative synthetic images of PRISMA's TANGO spacecraft made publicly available by Stanford's Space Rendezvous Laboratory (SLAB) and applicable to train and test different network architectures. One of the main advantages of CNNs over standard feature-based algorithms for relative pose estimation [7,17,18] is an increase in robustness under adverse illumination conditions, as well as a reduction in the computational complexity. Initially, *end-to-end* CNNs were exploited to map a 2D input image directly into a relative pose by means of learning complex non-linear functions [19–22]. However, since the pose accuracies of these end-to-end CNNs proved to be lower than the accuracies returned by common pose estimation solvers, especially in the estimation of the relative attitude [19], recent efforts investigated the capability of CNNs to perform keypoint localization prior to the actual pose estimation [23–26]. The output of these networks is a set of so called *heatmaps* around pre-trained features. The 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 [23]. Additionally, 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 [27]. In this context, the High-Resolution Net (HRNet) [28] already proved to be a reliable and accurate keypoint detector prior to pose estimation, due to its capability of maintaining a high-resolution representation of the heatmaps through the whole detection process.

To the best of the authors' knowledge, the reviewed implementations of CNNs feed solely the heatmap's peak location into the pose estimation solver, despite multiple information could be extracted from the detected heatmaps. Only in Pavlakos et al. [23], the pose estimation is solved by assigning weights to each feature based on their heatmap's peak intensities, in order to penalize inaccurate detections. Yet, there is another aspect related to the heatmaps which has not been considered. It is in fact hardly acknowledged how the overall shape of the detected heatmaps returned by CNN can be translated into a statistical distribution around the peak, allowing reliable feature covariances and, in turn, a robust navigation performance. As already investigated by the authors in earlier works [29,30], deriving an accurate representation of the measurements uncertainty from feature heatmaps can in fact not only improve the pose estimation, but it can also benefit the estimation of the full relative state vector, which would include the relative pose as well as the relative translational and rotational velocities.

From a high level perspective, two different navigation architectures are normally exploited in the framework of relative pose estimation. A *tightly-coupled* architecture, where the extracted features are directly processed by the navigation filter as measurements, and a *loosely-coupled* architecture, in which the relative pose is computed by a pose solver prior to the navigation filter, in order to derive pseudomeasurements from the target features [31]. Usually, a loosely-coupled approach is preferred for an uncooperative tumbling target, due to the fact that the fast relative dynamics could jeopardize feature tracking and return highly-variable measurements to the filter. However, one shortcoming of this approach is that it is generally hard to obtain a representative covariance matrix for the pseudomeasurements. This can be quite challenging when filter robustness is demanded. Remarkably, the adoption of a CNN in the feature detection step can overcome the challenges in feature tracking by guaranteeing the detection of a constant, pre-defined set of features. At the same time, the CNN heatmaps can be used to derive a measurements covariance matrix and improve filter robustness. Following this line of reasoning, a tightly-coupled filter is expected to interface well with a CNN-based IP and to outperform its loosely-coupled counterpart.

In this framework, the objective of this paper is to combine a CNN-based feature detector with a CEPPnP solver whilst evaluating the performance of a proposed tightly-coupled navigation filter against the performance of a loosely-coupled filter. Specifically, the novelty of this work stands in extending the authors' previous findings [29,30] by further linking the current research on CNN-based feature detection, covariant-based PnP solvers, and navigation filters. The main contributions of this work are:

1. To assess the feasibility of a simplified CNN for feature detection within the IP
2. To improve the pose estimation by incorporating heatmaps-derived covariance matrices in the CEPPnP
3. To compare the performance of tightly- and loosely-coupled navigation filters.

The paper is organized as follows. The overall pose estimation framework is illustrated in Section 2. Section 3 introduces the proposed CNN architecture together with the adopted training, validation, and testing datasets. In Section 4, special focus is given to the derivation of covariance matrices from the CNN heatmaps, whereas Section 5 describes the CEPPnP solver. Besides, Section 6 provides a description of the tightly- and loosely-coupled filters adopted. The simulation environment is presented in Section 7 together with the simulation results. Finally, Section 8 provides the main conclusions and recommendations.

2. Pose estimation framework

This work considers a servicer spacecraft flying in relative motion around a target spacecraft located in a Low Earth Orbit (LEO), with the relative motion being described in a Local Vertical Local Horizontal (LVLH) reference frame co-moving with the servicer (Fig. 1a). 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 tied to each spacecraft's body. The distance between the origins of these two frames defines their relative position. Together, these two quantities characterize the relative pose. This information can then be transferred from the camera frame to the servicer's centre of mass by accounting for the relative pose of the camera with respect to the LVLH frame.

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 pose of such target with respect to the servicer camera. Referring to Fig. 1b, the pose estimation problem consists in determining the position of the target's centre of mass $\mathbf{t}^C$ and its orientation with respect to the camera frame C, represented by the rotation matrix $\mathbf{R}_B^C$. The Perspective-n-Points (PnP) equations,

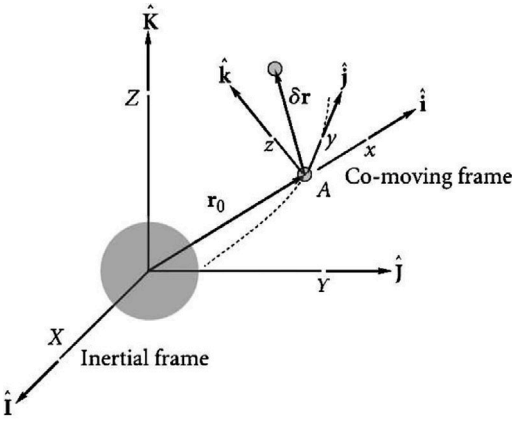
$$\mathbf{r}^C = \begin{pmatrix} x^C & y^C & z^C \end{pmatrix}^T = \mathbf{R}_B^C \mathbf{r}^B + \mathbf{t}^C \tag{1}$$

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

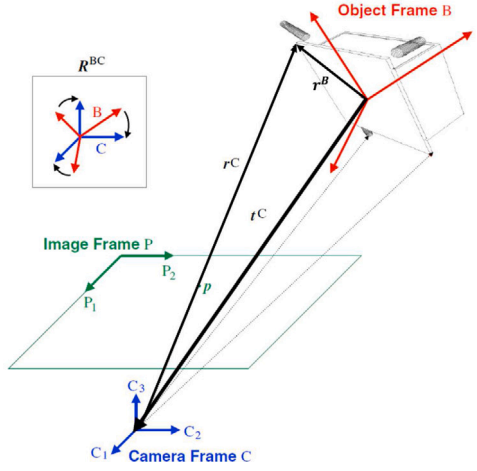
relate the unknown pose with a feature point $\mathbf{p}$ in the image plane via the relative position $\mathbf{r}^C$ of the feature with respect to the camera frame. Here, $\mathbf{r}^B$ is the point location in the 3D model, expressed in the body-frame coordinate system B, whereas f_x and f_y denote 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, such 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 $n = 3$ [33]. Generally, more features are required in presence of large noise and/or symmetric objects. Besides, it can also be expected that the time variation of the relative pose plays a crucial role while navigating around the target spacecraft, e.g. if rotational synchronization with the target spacecraft is required in the final approach phase. As such, it is clear that the estimation of both the relative translational and angular velocities represent an essential step within the navigation system.

The proposed tightly-coupled architecture combines the above key ingredients in three main stages, which are shown in Fig. 2 and described in more detail in the following sections. In the CNN-based IP block, a CNN is used to extract features from a 2D image of the target spacecraft. Statistical information is derived by computing a covariance matrix for each features using the information included in the output heatmaps. In the Navigation block, both the peak locations and the covariances are fed into the navigation filter, which estimates



(a) Co-moving LVLH frame [32]



(b) PnP problem (Figure adapted from [7])

Fig. 1. Representation of the relative motion framework (left) and schematic of the pose estimation problem using a monocular image (right) [32].

the relative pose as well as the relative translational and rotational velocities. The filter is initialized by the CEPPnP block, which takes peak location and covariance matrix of each feature as input and outputs the initial relative pose by solving the PnP problem in Eqs. (1)–(2). Thanks to the availability of a covariance matrix of the detected features, this architecture can guarantee a more accurate representation of feature uncertainties, especially in case of inaccurate detection of the CNN due to adverse illumination conditions and/or unfavourable relative geometries between servicer and target. Together with the CEPPnP initialization, this aspect can return a robust and accurate estimation of the relative pose and velocities and assure a safe approach of the target spacecraft.

In this work, a rectilinear VBAR approach of the servicer spacecraft towards the target spacecraft is considered, as this typically occurs during the final stages of close-proximity operations in rendezvous and docking missions [1,2]. This assumption is justified by the fact that the proposed method needs to be first validated on simplified relative trajectories before assessing its feasibility under more complex relative geometries. Following the same line of reasoning, the relative attitude is also simplified by considering a perturbation-free rotational dynamics between the servicer and the target. This is described in more detail in Section 6.

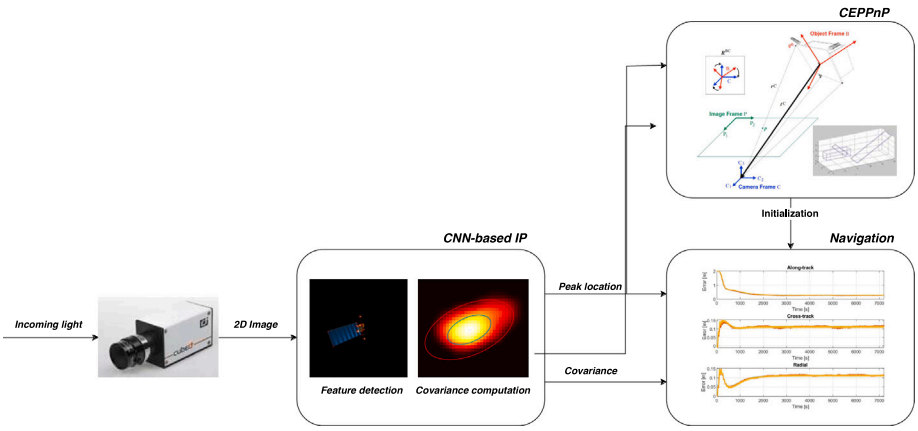


Fig. 2. Functional flow of the proposed tightly-coupled pose estimation architecture.

3. Convolutional neural network

CNNs are currently emerging as a promising features extraction method, mostly due to the capability of their convolutional layers to extract high-level features of objects with improved robustness against image noise and illumination conditions. In order to optimize CNNs for the features extraction process, a stacked hourglass architecture has been proposed [23,24], and other architectures such as the U-net [34] and the HRNet [28] were tested in recent years.

Compared to the network proposed by Pavlakos et al. [23], the architecture proposed in this work is composed of only one encoder/decoder block, constituting a single hourglass module. This was chosen in order to reduce the network size and comply with the limitations in computing power which characterizes space-grade processors. The encoder includes six blocks, each including a convolutional layer formed by a fixed number of filter kernels of size 3×3, a batch normalization module and max pooling layer, whereas the six decoder blocks accommodate an up-sampling block in spite of max pooling. In the encoder stage, the initial image resolution is decreased by a factor of two, with this downsampling process continuing until reaching the lowest resolution of 4×4 pixels. An upsampling process follows in the decoder with each layer increasing the resolution by a factor of two and returning output heatmaps at the same resolution as the input image. Fig. 3 shows the high-level architecture of the network layers, together with the corresponding input and output.

Overall, the size of the 2D input image and the number of kernels per convolutional layer drive the total number of parameters. In the current analysis, an input size of 256×256 pixels is chosen, and 128 kernels are considered per convolutional layer, leading to a total of $\sim$ 1,800,000 trainable parameters. Compared to the CNNs analysed by Sun et al. [28], this represents a reduction of more than an order of magnitude in network size.

As already mentioned, the output of the network is a set of heatmaps around the selected features. Ideally, the heatmap's peak intensity associated to a wrong detection should be relatively small compared to the correctly detected features, highlighting that the network is not confident about that particular wrongly-detected feature. At the same time, the heatmap's amplitude should provide an additional insight into the confidence level of each detection, a large amplitude being related to large uncertainty about the detection. The network is trained with the x- and y- image coordinates of the feature points, computed offline based on the intrinsic camera parameters as well as on the feature coordinates in the target body frame, which were extracted from the wireframe 3D model prior to the training. During training, the network is optimized to locate 16 features of the Envisat spacecraft, consisting of the corners of the main body, the Synthetic-Aperture Radar (SAR) antenna, and the solar panel, respectively. Fig. 4 illustrates the selected features for a specific target pose.

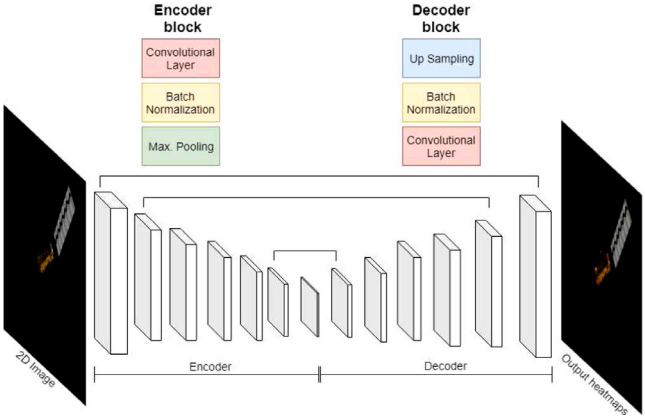


Fig. 3. Overview of the single hourglass architecture. Downsampling is performed in the *encoder* stage, in which the image size is decrease after each block, whereas upsampling occurs in the *decoder* stage. The output of the network consists of heatmap responses, and is used for keypoints localization.

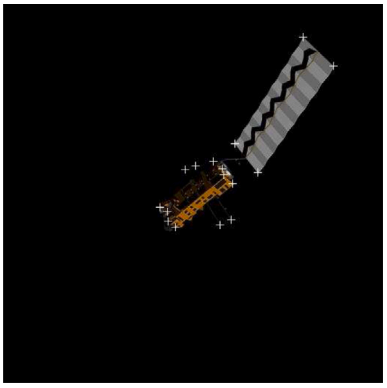


Fig. 4. Illustration of the selected features for a given Envisat pose.

3.1. Training, validation and test

For the training, validation, and test datasets, synthetic images of the Envisat spacecraft were rendered in the Cinema 4D©software. Table 1 lists the main camera parameters adopted. Constant Sun elevation and azimuth angles of 30 degrees were chosen in order to recreate favourable as well as adverse illumination conditions. Relative distances between camera and target were discretized every 30 m in the interval 90 m - 180 m, with the Envisat always located in the

Table 1
Parameters of the camera used to generate the synthetic images in Cinema 4D©.

Parameter	Value	Unit
Image resolution	512 × 512	pixels
Focal length	$3.9 \cdot 10^{-3}$	m
Pixel size	$1.1 \cdot 10^{-5}$	m

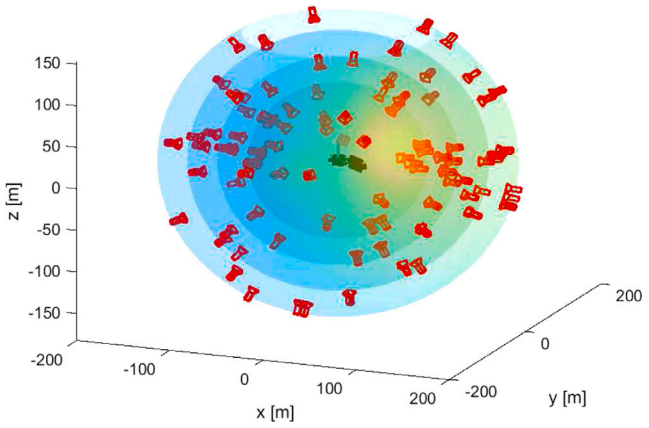


Fig. 5. Illustration of the pose space discretization in the training dataset. The concentric spheres represent the discretization of the relative distance in the range 90 m–180 m. Only 100 random relative camera poses are shown for clarity.

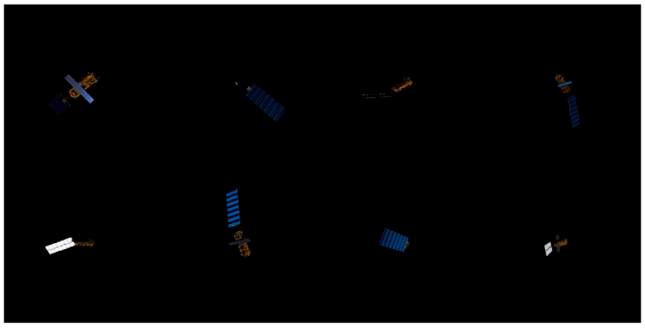


Fig. 6. A montage of eight synthetic images selected from the training set.

camera boresight direction in order to prevent some of the Envisat features from falling outside the camera field of view. Although being a conservative assumption, this allows to test the CNN detection under ideal servicer-target geometries during a rectilinear approach. Subsequently, relative attitudes were generated by discretizing the yaw, pitch, and roll angles of the target with respect to the camera by 10 degrees each. Together, these two choices were made in order to recreate several relative attitudes between the servicer and the target. The resulting database was then shuffled to randomize the images, and was ultimately split into training (18,000 images), validation (6,000 images), and test (6,000 images) datasets. Fig. 5 shows a subset of the camera pose distribution for 100 representative training images, whereas Fig. 6 illustrates some of the images included in the training dataset.

During training, the validation dataset is used beside the training one to compute the validation losses and avoid overfitting. The Adam optimizer [35] is used with a learning rate of 10^{-3} for a total number of 50 epochs. Finally, the network performance after training is assessed with the test dataset.

Preliminary results on the single-stack network performance were already reported by Pasqualetto Cassinis et al. [29]. Above all, one key advantage of relying on CNNs for feature detection was found in the capability of learning the relative position between features under

Table 2
Mean μ and Standard Deviation σ of the adopted networks over the Envisat test dataset.

Network	No. Params	μ [pxl]	σ [pxl]
Single-stack Hourglass	1.8M	3.4	4.3
HRNet	25M	2.4	1.4

a variety of relative poses present in the training. As a result, both features which are not visible due to adverse illumination and features occulted by other parts of the target can be detected. Besides, a challenge was identified in the specific selection of the trainable features. Since the features selected in this work represent highly symmetrical points of the Envisat spacecraft, such as corners of the solar panel, SAR antenna or main body, the network could be unable to distinguish between similar features, and return multiple heatmaps for a single feature output. Fig. 7 illustrates these findings. Notably, the detection of wrong features results in weak heatmaps, which can be filtered out by selecting a proper threshold on their total brightness.

In order to compare the feature detection accuracy of the proposed Single-stack Hourglass with a more complex CNN architecture, the HRNet proposed by Sun et al. [28] has been selected and trained on the same Envisat datasets. This architecture had already been tested on the SPEED dataset [25] and already proved to return highly accurate features of the TANGO spacecraft. The performance is assessed in terms of Root Mean Squared Error (RMSE) between the ground truth (GT) and the x, y coordinates of the extracted features, which is computed as

$$E_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^{n_{tot}} \left[(x_{GT,i} - x_i)^2 + (y_{GT,i} - y_i)^2\right]}{n_{tot}}}. \tag{3}$$

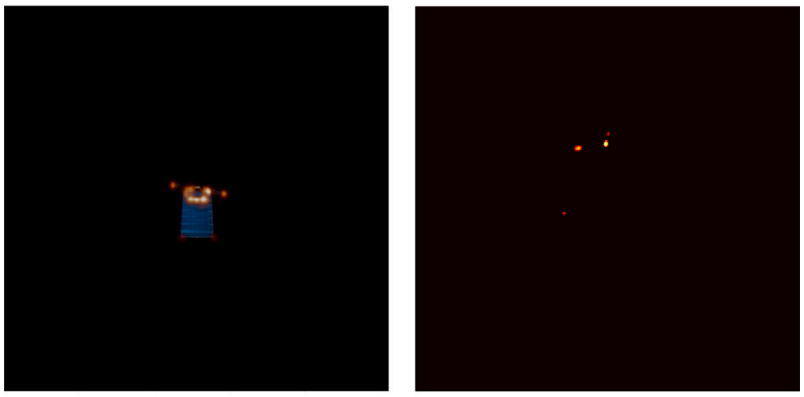
Fig. 8 shows the RMSE error over the test dataset for the two CNNs, whereas Table 2 reports the mean μ and standard deviation σ of the associated histograms. As expected, the added complexity of HRNet, translates into a more accurate detection of the selected features, thanks to the higher number of parameters: only 4% of the test images are characterized by a RMSE above 5 pixels, as opposed to the 15% in the Single-stack Hourglass case.

Although HRNet proves to return more accurate features, it is also believed that the larger RMSE scenarios returned by the Single-stack Hourglass can be properly handled, if a larger uncertainty can be associated to their corresponding heatmaps. As an example, a large RMSE could be associated to the inaccurate detection of only a few features which, if properly weighted, could not have a severe impact on the pose estimation step. This task can be performed by deriving a covariance matrix for each detected feature, in order to represent its detection uncertainty. Above all, this may prevent the pose solver and the navigation filter from trusting wrong detections by relying more on other accurate features. In this way, the navigation filter can cope with poorly accurate heatmaps while at the same time relying on a computationally-low CNN.

4. Covariance computation

Compared to the methods discussed in Section 1 [13–15], the proposed method derives a 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 order to do so, the first step is to obtain a statistical population around the heatmap's peak. This is done by thresholding each heatmap image so that only the x- and y- location of heatmap's pixels are extracted. Secondly, each pixel within the population is given a normalized weight w_i based on the grey intensity I_i at its location,

$$w_i = w_R R_i + w_G G_i + w_B B_i, \tag{4}$$



(a) Occulted spacecraft body (b) Heatmap of a single solar panel's corner

Fig. 7. Robustness and challenges of feature detection with the proposed CNN. On the left-hand side, the network has been trained to recognize the pattern of the features, and can correctly locate the body features which are not visible, i.e. parts occulted by the solar panel and corners of the SAR antenna. Conversely, the right-hand side shows the detection of multiple heatmaps for a single corner of the solar panel. As can be seen, the network can have difficulties in distinguishing similar features, such as the corners of the solar panel.

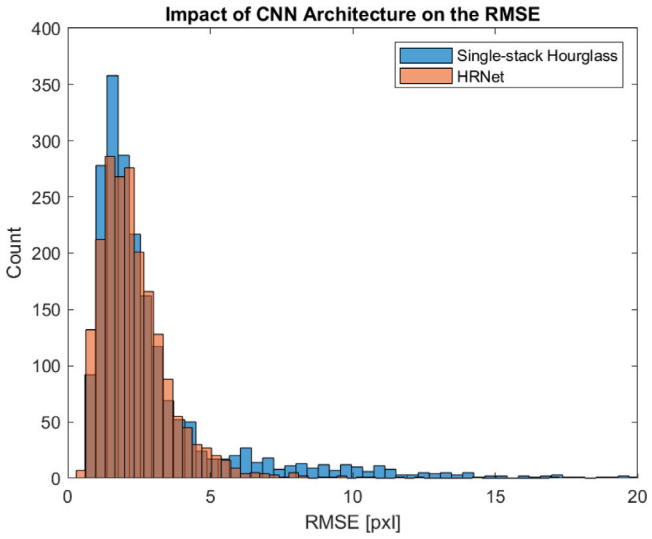


Fig. 8. RMSE over the test dataset for the HRNet and the Single-stack Hourglass.

where R, G, B are the components of the coloured image and w_R, w_G, w_B are the weights assigned to each channel in order to get the greyscale intensity. This is done in order to give more weight to pixels which are particularly bright and close to the peak, and less weight to pixels which are very faint and far from the peak. Finally, the obtained statistical population of each feature is used to compute the weighted covariance between x, y and consequently the covariance matrix $\boldsymbol{C}_i$,

$$\boldsymbol{C}_i = \begin{pmatrix} \text{cov}(x,x) & \text{cov}(x,y) \\ \text{cov}(y,x) & \text{cov}(y,y) \end{pmatrix}, \tag{5}$$

where

$$\text{cov}(x,y) = \sum_{i=1}^{n} w_i (x_i - p_x) \cdot (y_i - p_y) \tag{6}$$

and n is the number of pixels in each feature's heatmap. In this work, the mean is replaced by the peak location $\boldsymbol{p} = (p_x, p_y)$ in order to represent a distribution around the peak of the detected feature, rather than around the heatmap's mean. This is particularly relevant when the heatmaps are asymmetric and their mean does not coincide with their peak.

Fig. 9 shows the overall flow to obtain the covariance matrix for three different heatmap shapes. The ellipse associated to each features covariance is obtained by computing the eigenvalues λ_x and λ_y of the covariance matrix,

$$\left(\frac{x}{\lambda_x}\right)^2 + \left(\frac{y}{\lambda_y}\right)^2 = s, \tag{7}$$

where s defines the scale of the ellipse and is derived from the confidence interval of interest, e.g. $s = 2.2173$ for a 68% confidence interval. As can be seen, different heatmaps can result in very different covariance matrices. Above all, the computed covariance can capture the different CNN uncertainty over x, y. Notice that, due to its symmetric nature, the covariance matrix can only represent bivariate normal distributions. As a result, asymmetrical heatmaps such as the one in the third scenario are approximated by Gaussian distributions characterized by an ellipse which might overestimate the heatmap's dispersion over some directions.

5. Pose estimation

The CEPPnP method proposed by Ferraz et al. [13] was selected to estimate the relative pose from the detected features as well as from their covariance matrices. The first step of this method is to rewrite the PnP problem in Eqs. (1)–(2) as a function of a 12-dimensional vector $\boldsymbol{y}$ containing the control point coordinates in the camera reference system,

$$\boldsymbol{M}\boldsymbol{y} = \boldsymbol{0}, \tag{8}$$

where $\boldsymbol{M}$ is a $2n \times 12$ known matrix. This is the fundamental equation in the EPnP problem [9]. The likelihood of each observed feature location $\boldsymbol{u}_i$ is then represented as

$$P(\boldsymbol{u}_i) = k \cdot e^{-\frac{1}{2}\Delta\boldsymbol{u}_i^T \boldsymbol{C}_{\boldsymbol{u}_i}^{-1} \Delta\boldsymbol{u}_i}, \tag{9}$$

where $\Delta\boldsymbol{u}_i$ is a small, independent and unbiased noise with expectation $E[\Delta\boldsymbol{u}_i] = \boldsymbol{0}$ and covariance $E[\Delta\boldsymbol{u}_i \Delta\boldsymbol{u}_i^T] = \sigma^2 \boldsymbol{C}_{\boldsymbol{u}_i}$ and k is a normalization constant. Here, σ^2 represents the global uncertainty in the image, whereas $\boldsymbol{C}_{\boldsymbol{u}_i}$ is the 2×2 unnormalized covariance matrix representing the Gaussian distribution of each detected feature, computed from the CNN heatmaps. After some calculations [13], the EPnP formulation can be rewritten as

$$(\boldsymbol{N} - \boldsymbol{L})\boldsymbol{y} = \lambda \boldsymbol{y}. \tag{10}$$

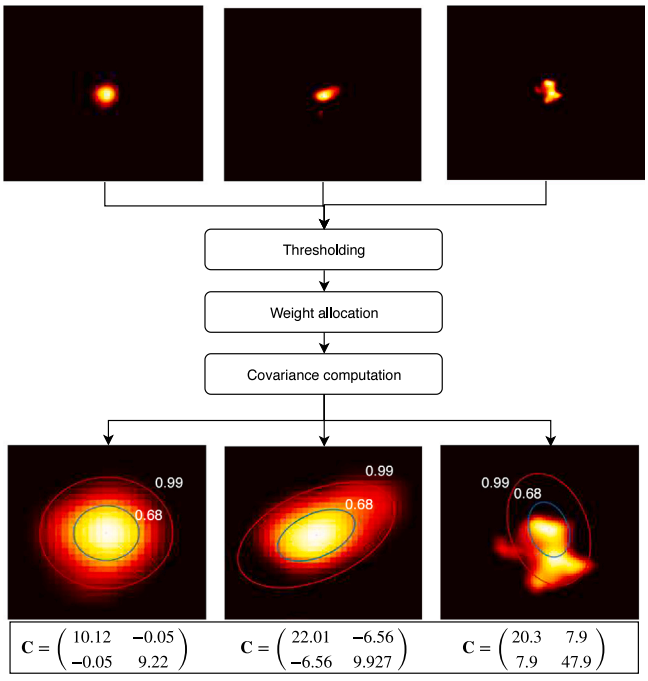


Fig. 9. Schematic of the procedure followed to derive covariance matrices from CNN heatmaps. The displayed ellipses are derived from the computed covariances by assuming the confidence intervals $\mathbf{1}\sigma = \mathbf{0.68}$ and $\mathbf{3}\sigma = \mathbf{0.99}$.

This is an eigenvalue problem in which both $\boldsymbol{N}$ and $\boldsymbol{L}$ matrices are a function of $\boldsymbol{y}$ and $\boldsymbol{C}_{u_i}$. The problem is solved iteratively by means of the closed-loop EPPnP solution for the four control points, assuming no feature uncertainty.

Once $\boldsymbol{y}$ is estimated, the relative pose is computed by solving the generalized Orthogonal Procrustes problem used in the EPPnP [10].

6. Navigation filter

Several navigation filters for close-proximity operations were investigated in recent years in the context of relative pose estimation. The reader is referred to Pasqualetto Cassinis et al. [8] for a comprehensive overview that goes beyond the scope of this work. In the proposed navigation system, the so-called Multiplicative Extended Kalman Filter (MEKF) is used. Remarkably, other works [15,30] adopted a standard formulation of the EKF that propagates the relative pose, expressed in terms of relative position and quaternions, as well as the relative translational and rotational velocities (prediction step), correcting the prediction with the measurements obtained from the monocular camera (correction step). However, the quaternion set consists of four parameters to describe the 3DOF attitude, hence one of its parameters is deterministic. As reported by Tweddle and Saenz-Otero [36] and Sharma and D'Amico [31], this makes the covariance matrix of a quaternion have one eigenvalue that is exactly zero. As a result, the entire state covariance propagated by the filter may become non-positive-definite and lead to the divergence of the filter. The MEKF, introduced for the first time by Lefferts et al. [37], 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.

6.1. Propagation step

A standard EKF state vector for relative pose estimation is composed of the relative pose between the servicer and the target, as well as the relative translational and rotational velocities $\boldsymbol{v}$ and $\boldsymbol{\omega}$. Under the assumption that the camera frame onboard the servicer is co-moving with the LVLH frame, with the camera boresight aligned with the along-track direction, this translates into

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{t}^C & \boldsymbol{v} & \boldsymbol{q} & \boldsymbol{\omega} \end{pmatrix}^T, \tag{11}$$

where $\boldsymbol{q} = \begin{pmatrix} q_0 & \boldsymbol{q_v} \end{pmatrix}$ is the quaternion set that represents the relative attitude. Notice that the assumption of the camera co-moving with the LVLH is made only to focus on the navigation aspects rather than on the attitude control of the servicer. Therefore, the application of the filter can be extended to other scenarios, if attitude control is included in the system.

In the MEKF, the modified state vector propagated inside the filter becomes

$$\tilde{\boldsymbol{x}} = \begin{pmatrix} \boldsymbol{t}^C & \boldsymbol{v} & \boldsymbol{a} & \boldsymbol{\omega} \end{pmatrix}^T, \tag{12}$$

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

$$\boldsymbol{a} = 4\boldsymbol{\sigma} = 4\frac{\boldsymbol{q_v}}{1+q_0}. \tag{13}$$

The discrete attitude propagation step is derived by linearizing $\dot{\boldsymbol{a}}$ around $\boldsymbol{a} = \boldsymbol{0}_{3\times1}$ and assuming small angle rotations [36],

$$\dot{\boldsymbol{a}} = \frac{1}{2}[\boldsymbol{\omega}\times]\boldsymbol{a} + \boldsymbol{\omega}. \tag{14}$$

As a result, the discrete linearized propagation of the full state becomes

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{\Phi}_k \tilde{\boldsymbol{x}}_{k-1} + \boldsymbol{\Gamma}_k \boldsymbol{Q}_k, \tag{15}$$

where $\boldsymbol{Q}_k$ represents the process noise and

$$\boldsymbol{\Phi}_k = \begin{pmatrix} \boldsymbol{\Phi}_{CW} & \boldsymbol{0}_{6\times6} \\ \boldsymbol{0}_{6\times6} & \boldsymbol{\Phi}_{a,\omega} \end{pmatrix} \tag{16}$$

$$\boldsymbol{\Phi}_{CW} = \begin{pmatrix} \boldsymbol{\Phi}_{rr} & \boldsymbol{\Phi}_{rv} \\ \boldsymbol{\Phi}_{vr} & \boldsymbol{\Phi}_{vv} \end{pmatrix} \tag{17}$$

$$\boldsymbol{\Phi}_{rr} = \begin{pmatrix} 1 & 0 & 6(\Delta\theta - \sin\Delta\theta) \\ 0 & \cos\Delta\theta & 0 \\ 0 & 0 & 4 - 3\cos\Delta\theta \end{pmatrix} \tag{18}$$

$$\boldsymbol{\Phi}_{rv} = \begin{pmatrix} 1/\omega_s(4\sin\Delta\theta - 3\Delta\theta) & 0 & 2/\omega_s(1-\cos\Delta\theta) \\ 0 & 1/\omega_s\sin\Delta\theta & 0 \\ 2/\omega_s(\cos\Delta\theta - 1) & 0 & \sin\Delta\theta/\omega \end{pmatrix} \tag{19}$$

$$\boldsymbol{\Phi}_{vr} = \begin{pmatrix} 0 & 0 & 6\omega_s(1-\cos\Delta\theta) \\ 0 & \omega_s\sin\Delta\theta & 0 \\ 0 & 0 & 3\omega_s\sin\Delta\theta \end{pmatrix} \tag{20}$$

$$\boldsymbol{\Phi}_{vv} = \begin{pmatrix} 4\cos\Delta\theta - 3 & 0 & 2\sin\Delta\theta \\ 0 & \cos\Delta\theta & 0 \\ -2\sin\Delta\theta & 0 & \cos\Delta\theta \end{pmatrix} \tag{21}$$

$$\boldsymbol{\Phi}_{a,\omega} = \begin{pmatrix} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\Delta t} & \int_0^{\Delta t} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\tau} d\tau \\ \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} \end{pmatrix} \tag{22}$$

$$\boldsymbol{\Gamma}_{k+1} = \begin{pmatrix} \frac{1}{2m}\boldsymbol{I}_{3\times3}\Delta t^2 & \boldsymbol{0}_{3\times3} \\ \frac{1}{m}\boldsymbol{I}_{3\times3}\Delta t & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \Delta t\int_0^{\Delta t} \boldsymbol{e}^{\frac{1}{2}[\omega\times]\tau}\boldsymbol{J}^{-1}d\tau \\ \boldsymbol{0}_{3\times3} & \boldsymbol{J}^{-1}\Delta t \end{pmatrix}. \tag{23}$$

The terms ω_s and $\Delta\theta$ in Eq. (17) represent the servicer argument of perigee and true anomaly variation from time t_0 to t, respectively, whereas the term $\boldsymbol{J}$ in Eq. (23) is the inertia matrix of the target spacecraft. In Tweddle and Saenz-Otero [36], the integral terms ins Eqs. (22)–(23) are solved by creating a temporary linear system from Eq. (14), augmented with the angular velocity and the process noise. The State Transition Matrix of this system is then solved numerically with the matrix exponential.

6.2. Correction step

At this stage, the propagated state $\tilde{\boldsymbol{x}}_k$ is corrected with the measurements $\boldsymbol{z}$ to return an update of the state $\hat{\boldsymbol{x}}_k$. In a loosely-coupled filter, these measurements are represented by the relative pose between the servicer and the target spacecraft, obtained by solving the PnP problem with the CEPPnP solver described in Section 5. In this case, a pseudomeasurements vector is derived by transforming the relative quaternion set into the desired attitude error $\boldsymbol{a}$,

$$\delta\boldsymbol{q}_z = \boldsymbol{q}_z \otimes \boldsymbol{q}_{\text{ref}_k}^* \rightarrow \boldsymbol{a} = 4\frac{\delta\boldsymbol{q}_v}{1+\delta q_0} \tag{24}$$

$$\boldsymbol{z}_k = \begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{a} \end{pmatrix} = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{V}_k = \begin{pmatrix} \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} \\ \boldsymbol{0}_{3\times3} & \boldsymbol{0}_{3\times3} & \boldsymbol{I}_{3\times3} & \boldsymbol{0}_{3\times3} \end{pmatrix} \begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \\ \boldsymbol{a} \\ \boldsymbol{\omega} \end{pmatrix}_k + \begin{pmatrix} \boldsymbol{V}_r \\ \boldsymbol{V}_a \end{pmatrix}_k. \tag{25}$$

In Eq. (24), $\otimes$ denotes the quaternion product. Conversely, in a tightly-coupled filter the measurements are represented by the pixel coordinates of the detected features,

$$\boldsymbol{z} = \begin{pmatrix} x_1, y_1 & \dots & x_n, y_n \end{pmatrix}^T. \tag{26}$$

Referring to Eqs. (1)–(2), this translates into the following equations for each detected point $\boldsymbol{p}_i$:

$$\boldsymbol{h}_i = \left(\frac{x_i^C}{z_i^C} f_x + C_x, \frac{y_i^C}{z_i^C} f_y + C_y \right)^T \tag{27}$$

$$\mathbf{r}^C = \boldsymbol{q} \otimes \boldsymbol{r}_i^B \otimes \boldsymbol{q}^* + \mathbf{t}^C, \tag{28}$$

where $\boldsymbol{q}^*$ is the quaternion conjugate. As a result, the measurements update equation can be written as

$$\boldsymbol{z}_k = \boldsymbol{H}_k\boldsymbol{x}_k + \boldsymbol{V} = \begin{pmatrix} \boldsymbol{H}_{t^C,i} & \boldsymbol{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{a},i} & \boldsymbol{0}_{2n\times3} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{H}_{t^C,n} & \boldsymbol{0}_{2n\times3} & \boldsymbol{H}_{\boldsymbol{a},n} & \boldsymbol{0}_{2n\times3} \end{pmatrix} \begin{pmatrix} \boldsymbol{t}^C \\ \boldsymbol{v} \\ \boldsymbol{a} \\ \boldsymbol{w} \end{pmatrix}_k + \begin{pmatrix} \boldsymbol{V}_r \\ \boldsymbol{V}_a \end{pmatrix} \tag{29}$$

and the Jacobian $\boldsymbol{H}_k$ of the observation model with respect of the state vector is a $2n\times13$ matrix whose elements are

$$\boldsymbol{H}_{r,i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{t^C,i}^{\text{ext}} \tag{30}$$

$$\boldsymbol{H}_{\boldsymbol{a},i} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{a},i}^{\text{ext}} = \boldsymbol{H}_i^{\text{int}} \cdot \boldsymbol{H}_{\boldsymbol{q},i}^{\text{ext}} \cdot \boldsymbol{H}_{\boldsymbol{a}}^{\boldsymbol{q}} \tag{31}$$

$$\boldsymbol{H}_i^{\text{int}} = \frac{\partial \boldsymbol{h}_i}{\partial \boldsymbol{r}_i^C} = \begin{pmatrix} \frac{f_x}{z_i^C} & 0 & -\frac{f_x}{(z_i^C)^2} x_i^C \\ 0 & \frac{f_y}{z_i^C} & -\frac{f_y}{(z_i^C)^2} y_i^C \end{pmatrix} \tag{32}$$

$$\boldsymbol{H}_{\boldsymbol{q},i}^{\text{ext}} = \frac{\partial \boldsymbol{r}_i^C}{\partial \boldsymbol{q}} = \frac{\partial\left(\boldsymbol{q} \otimes \boldsymbol{r}_i^B \otimes \boldsymbol{q}^*\right)}{\partial \boldsymbol{q}}; \quad \boldsymbol{H}_{t^C,i}^{\text{ext}} = \frac{\partial \boldsymbol{r}_i^C}{\partial \boldsymbol{t}^C} = \boldsymbol{I}_3 \tag{33}$$

$$\boldsymbol{H}_{\boldsymbol{a}}^{\boldsymbol{q}} = \frac{\partial\left(\delta\boldsymbol{q} \otimes \boldsymbol{q}_{\text{ref}}\right)}{\partial \boldsymbol{a}} = \frac{\partial\left(\boldsymbol{Q}_{\text{ref}}\delta\boldsymbol{q}\right)}{\partial \boldsymbol{a}} = \boldsymbol{Q}_{\text{ref}} \frac{\partial(\delta\boldsymbol{q})}{\partial \boldsymbol{a}} \tag{34}$$

$$\boldsymbol{Q}_{\text{ref}} = \begin{pmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{pmatrix}_{\text{ref}}. \tag{35}$$

The partial derivatives of the differential quaternion set with respect to the attitude error are computed from the relation between the attitude error $\boldsymbol{a}$ and the differential quaternion set $\delta\boldsymbol{q}$,

$$\delta q_0 = \frac{16 - \|\boldsymbol{a}\|^2}{16 + \|\boldsymbol{a}\|^2} \quad \delta\boldsymbol{q}_{\boldsymbol{v}} = 8\frac{\boldsymbol{a}}{16 + \|\boldsymbol{a}\|^2} \tag{36}$$

$$\frac{\partial(\delta\boldsymbol{q})}{\partial \boldsymbol{a}} = \frac{8}{\left(16 + \|\boldsymbol{a}\|^2\right)^2} \begin{pmatrix} -8a_1 & -8a_2 & -8a_3 \\ 16 + \|\boldsymbol{a}\|^2 - 2a_1^2 & -2a_1a_2 & -2a_1a_3 \\ -2a_2a_1 & 16 + \|\boldsymbol{a}\|^2 - 2a_2^2 & -2a_2a_3 \\ -2a_3a_1 & -2a_3a_2 & 16 + \|\boldsymbol{a}\|^2 - 2a_3^2 \end{pmatrix}. \tag{37}$$

In the tightly-coupled filter, the measurement covariance matrix $\boldsymbol{R}$ is a time-varying block diagonal matrix constructed with the heatmaps-derived covariances $\boldsymbol{C}_i$ in Eq. (5),

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{C}_1 & & \\ & \ddots & \\ & & \boldsymbol{C}_n \end{pmatrix}. \tag{38}$$

Notice that $\boldsymbol{C}_i$ can differ for each feature in a given frame as well as vary over time. Preliminary navigation results [30] already showed that such heatmaps-derived covariance matrix can capture the statistical distribution of the measured features and improve the measurements update step of the navigation filter. Conversely, in the loosely-coupled filter $\boldsymbol{R}$ represents the uncertainty in the pose estimation step and hence it is not directly related to the CNN heatmaps. A constant value is therefore chosen based on the pose estimation accuracy observed for the test dataset.

Finally, the updated state estimate $\hat{\boldsymbol{x}}_k$ is obtained from the propagated state $\tilde{\boldsymbol{x}}_k$, the residuals $\tilde{\boldsymbol{y}}$, and the Kalman Gain $\boldsymbol{K}$,

$$\tilde{\boldsymbol{y}} = \boldsymbol{z} - \boldsymbol{h}(\tilde{\boldsymbol{x}}_k) \tag{39}$$

$$\boldsymbol{K} = \boldsymbol{P}_k\boldsymbol{H}_k^T(\boldsymbol{H}_k\boldsymbol{P}_k\boldsymbol{H}_k^T + \boldsymbol{R}_k)^{-1} \tag{40}$$

$$\hat{\boldsymbol{x}}_k = \tilde{\boldsymbol{x}}_k + \boldsymbol{K}\tilde{\boldsymbol{y}}. \tag{41}$$

6.3. Reset step

In the reset step, the reference quaternion $\boldsymbol{q}_{\text{ref}}$ is updated with the attitude error estimate $\hat{\boldsymbol{a}}_p$ and the new attitude error is set to zero,

$$\hat{\boldsymbol{q}}_k = \delta\boldsymbol{q}(\hat{\boldsymbol{a}}) \otimes \boldsymbol{q}_{\text{ref}_k} \tag{42}$$

$$\hat{\boldsymbol{a}} = \boldsymbol{0}_{3\times1} \tag{43}$$

$$\boldsymbol{q}_{\text{ref}_{k+1}} = \hat{\boldsymbol{q}}_k. \tag{44}$$

The obtained estimated quaternion set $\hat{\boldsymbol{q}}_k$ is then compared to the real quaternion set to assess the angle accuracy of the filter.

7. Simulations

In this section, the simulation environment and the results are presented. Firstly, the impact of including a heatmaps-derived covariance in the pose estimation step is addressed by comparing the CEPPnP method with a standard solver which does not account for feature uncertainty. The weights in Eq. (4) are selected based on the standard RGB-to-greyscale conversion ($w_R = 0.299$, $w_G = 0.587$, $w_B = 0.114$). Secondly, the performance of the MEKF is evaluated by comparing the convergence profiles with a heatmaps-derived covariance matrix against covariance matrices with arbitrary selected covariances. Initialization is provided by the CEPPnP for all the scenarios.

Two separate error metrics are adopted in the evaluation, in accordance with Sharma and D'Amico [20]. Firstly, the translational error between the estimated relative position $\hat{\boldsymbol{t}}^C$ and the ground truth $\boldsymbol{t}$ is computed as

$$E_T = |\boldsymbol{t}^C - \hat{\boldsymbol{t}}^C|. \tag{45}$$

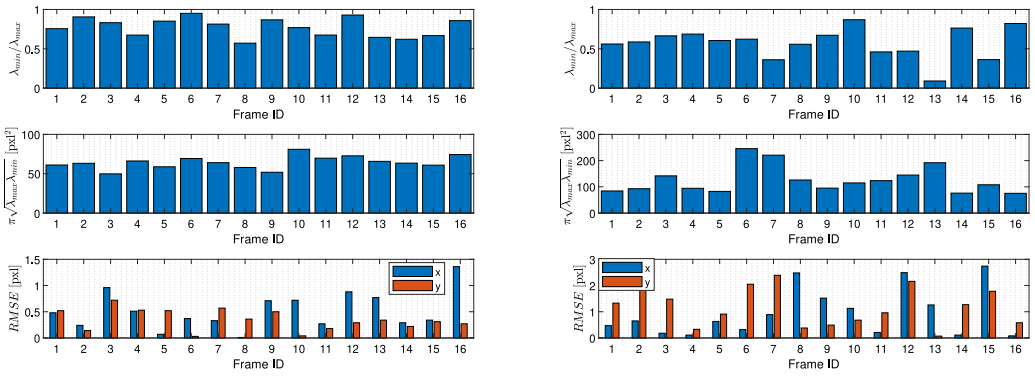
This metric is also applied for the translational and rotational velocities estimated in the navigation filter. Secondly, the attitude accuracy is measured in terms of the Euler axis–angle error between the estimated quaternion $\hat{\boldsymbol{q}}$ and the ground truth $\boldsymbol{q}$,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_s & \boldsymbol{\beta}_{\boldsymbol{v}} \end{pmatrix} = \boldsymbol{q} \otimes \hat{\boldsymbol{q}} \tag{46}$$

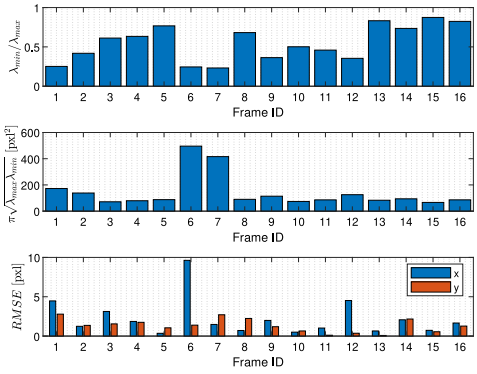
$$E_R = 2\arccos(|\beta_s|). \tag{47}$$

7.1. Pose estimation

Three representative scenarios are selected from the CNN test dataset for a preliminary evaluation of the Single-stack Hourglass

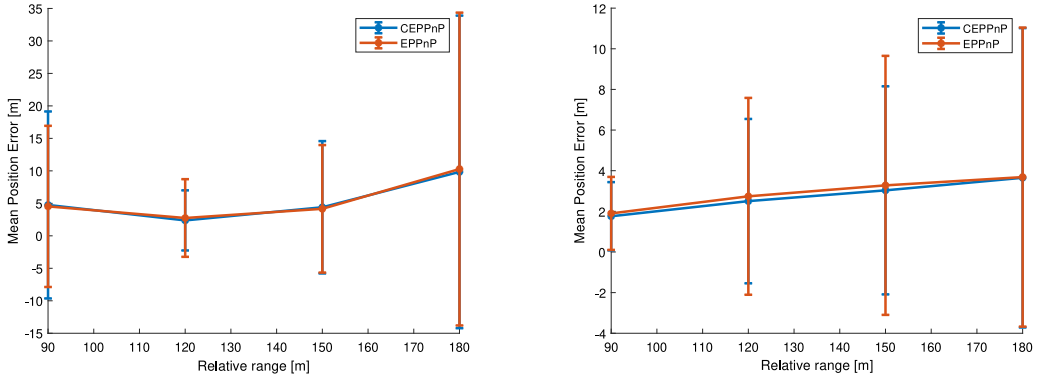


(a) Test scenario 1 (b) Test scenario 2



(c) Test scenario 3

Fig. 10. Characteristics of the ellipses derived from the covariance matrices for the three selected scenarios.



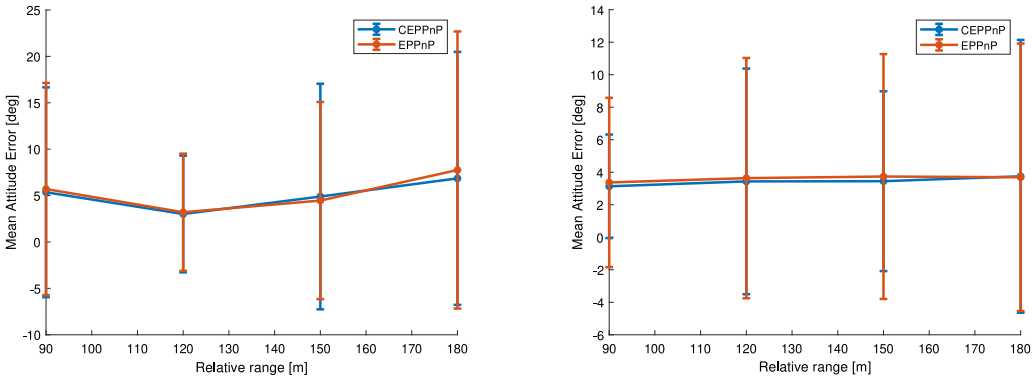
(a) Single-stack Hourglass (b) HRNet

Fig. 11. Pose Estimation Results - The standard deviation of the position error E_T is depicted as the length of each error bar above and below the mean error E_T.

performance. These scenarios were chosen in order to analyse different heatmaps' distributions around the detected features. A comparison is made between the proposed CEPPnP and the EPPnP. Fig. 10 shows the characteristics of the covariance matrices derived from the predicted heatmaps. Here, the ratio between the minimum and maximum eigenvalues of the associated covariances is represented against the ellipse's area and the RMSE between the Ground Truth (GT) and the x, y coordinates of the extracted features,

$$E_{\text{RMSE},i} = \sqrt{(x_{GT,i} - x_i)^2 + (y_{GT,i} - y_i)^2}. \tag{48}$$

Notably, interesting relations can be established between the three quantities reported in the figure. In the first scenario, the correlation between the sub-pixel RMSE and the large eigenvalues ratio suggests that a very accurate CNN detection can be associated with circular-shaped heatmaps. Moreover, the relatively low ellipse's areas indicate that, in general, small heatmaps are expected for an accurate detection. Conversely, in the second scenario the larger ellipses' area correlates with a larger RMSE. Furthermore, it can be seen that the largest difference between the x- and y- components of the RMSE occurs either for



(a) Single-stack Hourglass (b) HRNet

Fig. 12. Pose Estimation Results - The standard deviation of the attitude error E_R is depicted as the length of each error bar above and below the mean error E_R.

Table 3
Single-stack Hourglass Pose Estimation performance results for the selected three representative scenarios.

Metric	Scenario	CEPPnP	EPPnP
E_T [m]	1	[0.18 0.22 0.24]	[0.17 0.22 0.24]
	2	[0.35 0.41 0.59]	[0.14 0.4 22.8]
	3	[0.49 0.12 1.41]	[0.56 0.16 5.01]
E_R [deg]	1	0.36	0.35
	2	0.75	6.08
	3	1.99	2.72

Table 4
VBAR approach scenario. The attitude is represented in terms of ZYX Euler angles for clarity. Note that the camera boresight is the Y-axis of the LVLH frame.

$\boldsymbol{\theta}_0$ [deg]	$\boldsymbol{\omega}_0$ [deg/s]	$\boldsymbol{t}_0^C$ [m]	$\boldsymbol{v}_0$ [mm/s]
$[-180\ 30\ -80]^T$	$[-2.5\ -4.3\ 0.75]^T$	$[0\ 150\ 0]^T$	$[0\ 0\ 0]^T$

the most eccentric heatmap (ID 13) or for the one with the largest area (ID 6). The same behaviour can be observed in the last scenario, where the largest RMSE coincides with a large, highly eccentric heatmap.

Table 3 lists the pose estimation results for the three scenarios. As anticipated in Fig. 10, the statistical information derived from the heatmaps in the first scenario is uniform for all the features, due to the very accurate CNN detection. As a result, the inclusion of features covariance in the CEPPnP solver does not help refining the estimated pose. Both solvers are characterized by the same pose accuracy.

Not surprisingly, the situation changes as soon as the heatmaps are not uniform across the feature IDs. Due to its capability of accommodating feature uncertainties in the estimation, the CEPPnP method outperforms the EPPnP for the remaining scenarios. In other words, the CEPPnP solver proves to be more robust against inaccurate CNN detections by accounting for a reliable representation of the features covariance.

Next, the previous comparison is extended to the entire test dataset as well as to HRNet, by computing the mean and standard deviation of the estimated relative position and attitude as a function of the relative range, respectively. This is represented in Figs. 11–12. First of all, it can be seen that the pose accuracy of the CEPPnP solver in the Single-stack Hourglass scenario does not improve compared to the EPPnP, as opposed to the ideal behaviour reported in Table 3. There are two potential causes of this behaviour. On the one hand, most of the test images characterized by a large RMSE (Fig. 8) could not return statistically-meaningful heatmaps that would help the CEPPnP solver. This could be due to multiple heatmaps or highly inaccurate detections in which two different corners are confused with each other. On the other hand, this could be a direct consequence of the large relative ranges considered in this work. As already reported by Park et al. [26] and Sharma and D'Amico [31], a decreasing performance of EPPnP is indeed expected for increasing relative distances, due to the nonlinear relation between the pixel location of the detected features and z^C in Eq. (2). In other words, relatively large pixel errors could lead to inaccurate pose estimates for large relative distances, independently of the use of either CEPPnP or EPPnP.

Furthermore, it can be seen from a different comparison level that both the mean and standard deviation of the estimated relative pose are improved, when HRNet is used prior to the PnP solver (Figs. 11b-12b). Again, this is a direct consequence of the smaller RMSE reported in Fig. 8. As a result, the above-mentioned degradation of the pose estimation accuracy for increasing relative ranges is less critical for HRNet. Notice also that, despite an actual improvement of CEPPnP over EPPnP can be seen in the HRNet scenario, the improvements in both the mean and standard deviation of the estimation error are relatively small at large relative distances. This is considered to be related to the fact that HRNet returns circular heatmaps for most of the detected features, due to its higher detection accuracy compared to the Single-stack Hourglass.

Notably, it is important to assess how well the pose estimation system can scale when tested on datasets different than the Envisat one. To this aim, the proposed heatmaps-based scheme was benchmarked on the SPEED dataset, in order to compare its pose accuracy against standard as well as CNN-based systems [19,25,26]. The reader is referred to Barad [38, p. 115] for a comprehensive quantitative analysis of such comparison. The results demonstrated that the performance of the proposed pipeline, based on extracting feature heatmaps and using the CEPPnP solver, compares well with the state-of-the-art pose estimation systems.

7.2. Navigation filter

To assess the performance of the proposed MEKF, a rendezvous scenarios with Envisat is rendered in Cinema 4D©. This is a perturbation-free VBAR trajectory characterized by a relative velocity $\|\boldsymbol{v}\| = 0$ m/s. The Envisat performs a roll rotation of $\|\boldsymbol{\omega}\| = 5$ deg/s, with the servicer camera frame aligned with the LVLH frame. Table 4 lists the initial conditions of the trajectory, whereas Fig. 13 shows some of the associated rendered 2D images. It is assumed that the images are made available to the filter every 2 s for the measurement update step, with the propagation step running at 1 Hz. In both scenarios, the MEKF is initialized with the CEPPnP pose solution at time t_0. The other elements of the initial state vector are randomly chosen assuming a standard deviation of 1 mm/s and 1 deg/s for all the axes of terms $(\hat{\boldsymbol{v}}_0 - \boldsymbol{v})$ and $(\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega})$, respectively. Table 5 reports the initial conditions of the filter.

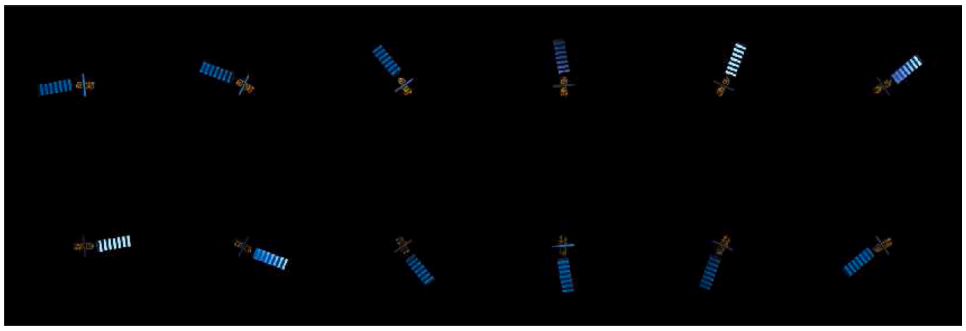


Fig. 13. Montage of the selected VBAR approach scenario. Images are shown every 6 s for clarity.

Table 5
Initial state vector in the MEKF. Here, HG refers to the Single-stack Hourglass architecture.

CNN	$\hat{\theta}_0$ [deg]	$\hat{\omega}_0$ [deg/s]	$\hat{t}_0^C$ [m]	$\hat{v}_0$ [mm/s]
HG	$[-180.2\ 28.7\ -80.6]^T$	$[-2.1\ -4.1\ 0.1]^T$	$[0.1\ 149.7\ 0.1]^T$	$[2.8\ -1.3\ 3]^T$
HRNet	$[-179.5\ 33.5\ -79.7]^T$	$[-2.1\ -4.1\ 0.1]^T$	$[-0.1\ 149.8\ -0.1]^T$	$[2.8\ -1.3\ 3]^T$

Table 6
Standard deviation of Monte Carlo variables.

$\sigma_{\Delta\phi_0}$ [deg]	σ_{ω_0} [deg/s]	$\sigma_{t_0^C}$ [m]	σ_{v_0} [mm/s]
10	1	$[1\ 10\ 1]^T$	10

Figs. 14–15 show the convergence profiles for the translational and rotational states in the tightly- and loosely-coupled MEKF, respectively. Besides, a Monte Carlo simulation with 1,000 runs was performed to assess the robustness of the filter estimate against varying the initial state $\hat{\boldsymbol{x}}_0$. Table 6 lists the standard deviation chosen for the deviation from the true initial state of the filter. The distribution follows a Gaussian profile with true-state mean. For the attitude initial error, the initial reference quaternion $\boldsymbol{q}_{\text{ref}_0}$ is perturbed by introducing a random angular error around the correct Euler axis [39, p. 44],

$$\Delta\boldsymbol{\phi}_0 = \Delta\phi_0 \boldsymbol{q}_{\boldsymbol{v}} \tag{49}$$

$$\boldsymbol{q}_{\text{ref}_0} = \boldsymbol{q}_0 \otimes \begin{pmatrix} 1 \\ \frac{1}{2}\Delta\boldsymbol{\phi}_0 \end{pmatrix}. \tag{50}$$

Table 7 reports the mean of the steady-state pose estimates together with their standard deviation. From these results, important insights can be gained on two different levels of the comparison.

On a CNN performance level, the results in Fig. 14 show that a slightly worse cross-track estimate of the Single-stack Hourglass is compensated by a more accurate estimate of the relative attitude. Given the limited impact of these estimation errors at the relatively large inter-satellite range of 150 m, these results suggest that the Single-stack Hourglass has a comparable performance with the HRNet for the selected scenario. Next, on a filter architecture level, a comparison between Figs. 14–15 illustrate the different convergence pattern between the tightly- and loosely-coupled MEKF. Most importantly, it can be seen that the loosely-coupled estimate of the relative along-track position is characterized by a bias which is not present in the tightly-coupled estimate. This occurs due to the decoupling of the translational and rotational states, reflected in the Jacobian $\boldsymbol{H}_k$ in Eq. (25). As a result, the relative position is estimated without accounting for the attitude measurements and vice versa. In other words, the creation of pseudomeasurements of the relative pose prior to the loosely-coupled filter leads to two separate translational and rotational estimates. Conversely, in the tightly-coupled filter the full statistical information is enclosed in the detected features, and can be used to simultaneously refine both the translational and the rotational states. Moreover, a close inspection of the Single-stack Hourglass attitude estimates in Table 7 suggests that the tightly-coupled MEKF is characterized by a lower standard deviation, highlighting a better robustness with respect to the initial conditions of the filter when compared to the loosely-coupled MEKF. Note that, due to the higher accuracy of HRNet in the feature detection step — and hence also in the pose estimation step, this is not observed for the latter CNN. In conclusion, a tightly-coupled architecture is expected to return higher pose accuracies if simplified CNNs, such as the proposed single-stack hourglass, are implemented at a feature detection level.

8. Conclusions and recommendations

This paper introduces a novel framework to estimate the relative pose of an uncooperative target spacecraft with a single monocular camera onboard a servicer spacecraft. A method is proposed in which a CNN-based IP algorithm is combined with a CEPPnP solver and a tightly-coupled MEKF to return a robust estimate of the relative pose as well as of the relative translational and rotational velocities. The performance of the proposed method is evaluated at different levels of the pose estimation system, by comparing the detection accuracy of two different CNNs (feature detection step and pose estimation step) whilst assessing the accuracy and robustness of the selected tightly-coupled filter against a loosely-coupled filter (navigation filter step).

The main novelty of the proposed CNN-based pose estimation system is to introduce a heatmaps-derived covariance representation of the detected features and to exploit this information in a tightly-coupled, Single-stack Hourglass-based MEKF. On a feature detection level, the performance of the proposed Single-stack Hourglass is compared to the more complex HRNet to assess the feasibility of a reduced-parameters CNN within the IP. Results on the selected test dataset suggest a comparable mean detection accuracy, despite a larger standard deviation of the former network. Notably, this latter aspect is found to decrease the pose estimation accuracy of the proposed CNN compared to HRNet, despite the adoption of CEPPnP to capture features uncertainty. However, important insights are gained at a navigation filter level, delineating two major benefits of the proposed tightly-coupled MEKF. First of all, the capability of deriving a measurements covariance matrix directly from the CNN heatmaps allows to capture a more representative statistical distribution of the measurements in the filter. Notably, this is expected to be a more complex task if a loosely-coupled filter is used, due to the need to convert the heatmaps distribution into a pose estimation uncertainty through a linear transformation. Secondly, the coupling between the rotational and translational states within the filter guarantees a mutual interaction which is expected to improve the global accuracy of the filter, especially in the along-track estimate. Besides, the navigation results for the selected VBAR

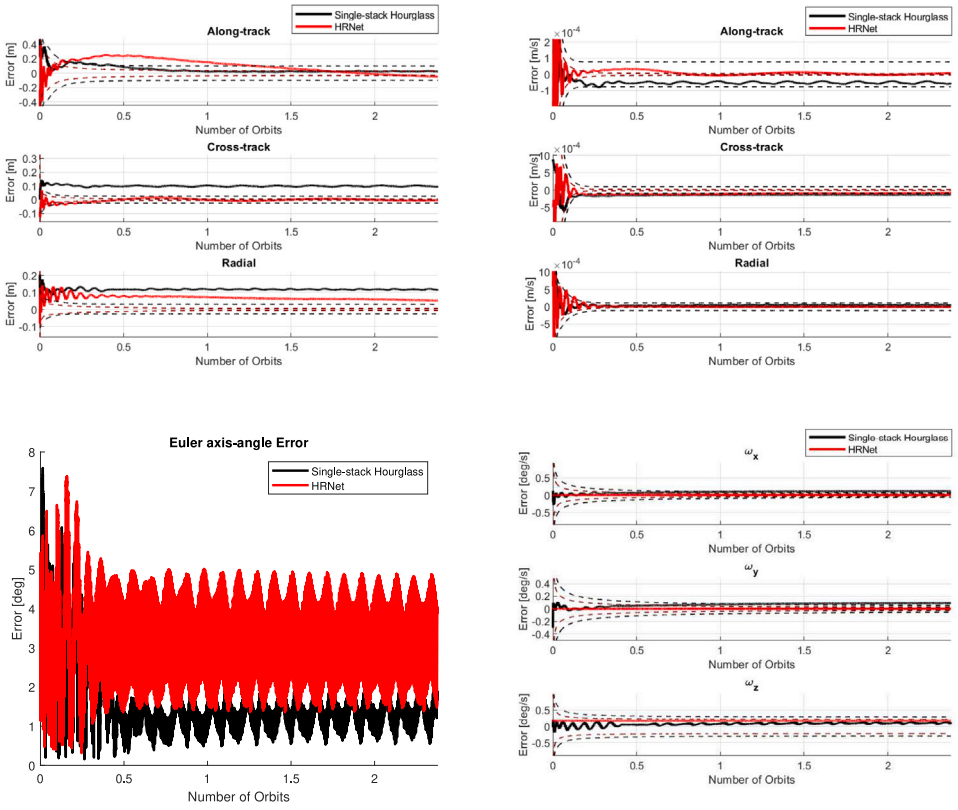


Fig. 14. Navigation Filter Results — Tightly-coupled MEKF. The dashed lines represent the 1σ of the estimated quantities.

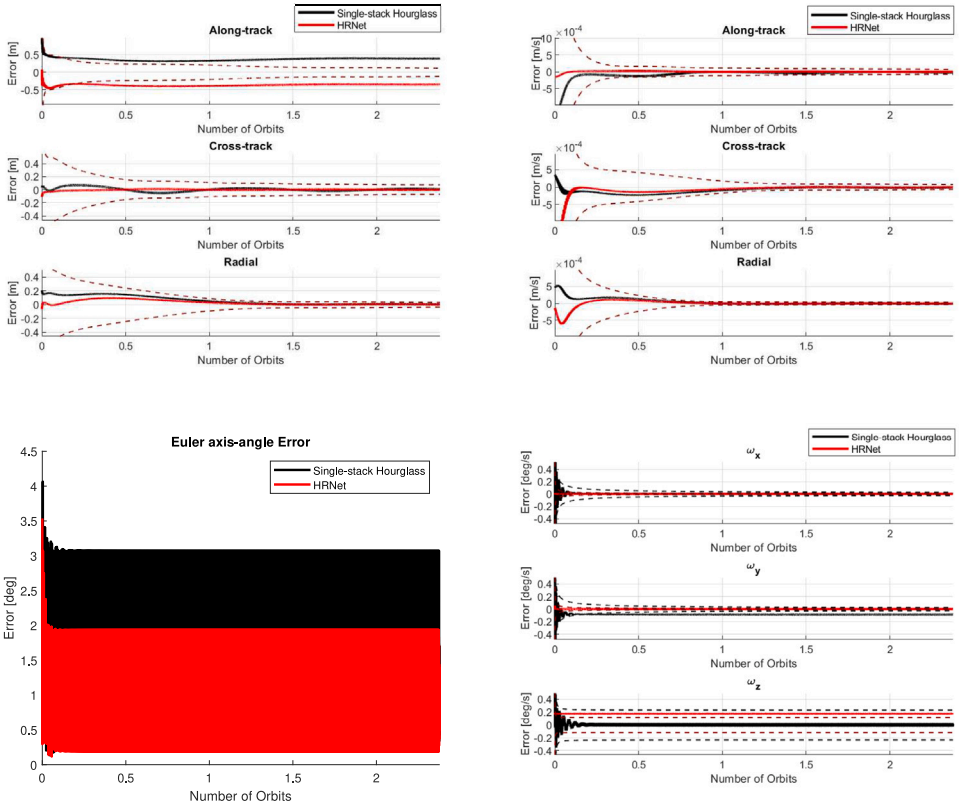


Fig. 15. Navigation Filter Results — Loosely-coupled MEKF. The dashed lines represent the 1σ of the estimated quantities.

Table 7
Monte Carlo Simulation Results. The mean and standard deviation of the relative pose errors are taken from the absolute errors at filter steady state for each Monte Carlo run.

Single-stack Hourglass				
MEKF	E_{T_1} [m]	E_{T_2} [m]	E_{T_3} [m]	E_R [deg]
Tightly-coupled	0.1182 ± 6.3E−4	0.03 ± 0.0024	0.096 ± 5.4E−4	1.33 ± 0.03
Loosely-coupled	0.1 ± 2E−7	0.33 ± 3E−7	0.01 ± 1E−4	4.7 ± 12.6
HRNet				
MEKF	E_{T_1} [m]	E_{T_2} [m]	E_{T_3} [m]	E_R [deg]
Tightly-coupled	0.0683 ± 5.2E−4	0.03 ± 0.014	0.01 ± 3.4E−5	4.3 ± 0.03
Loosely-coupled	0.0075 ± 1E−4	0.36 ± 6.3E−5	0.002 ± 4E−4	0.93 ± 3.2E−7

scenario demonstrated that the proposed Single-stack Hourglass could represent a valid alternative to the more complex HRNet, provided that its larger detection uncertainty is reflected in the measurements covariance matrix. Together, these improvements suggest a promising scheme to cope with the challenging demand for robust navigation in close-proximity scenarios.

However, further work is required in several directions. First of all, more recent CNN architectures shall be investigated to assess the achievable robustness and accuracy in the feature detection step. Secondly, the impact of a reduction in the number of CNN parameter on the computational complexity shall be assessed by testing the CNNs in space-representative processors. Moreover, broader relative ranges between the servicer camera and the target spacecraft shall be considered, most importantly to allow a thorough investigation of the 3D depth perception challenges when approaching the target spacecraft with a single monocular camera. Besides, more close-proximity scenarios shall be recreated to assess the impact of perturbations on the accuracy and robustness of the navigation filter. In this context, other navigation filters such as the Unscented Kalman Filter shall be investigated.

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.

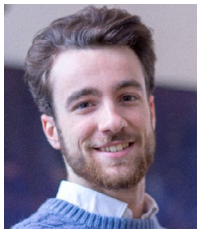
Acknowledgements

This study is funded and supported by the European Space Agency and Airbus Defence and Space under Network/Partnering Initiative (NPI) program with grant number NPI 577–2017. The first author would like to thank Christoph Haskampf for the rendering of the images in Cinema 4D©and Kuldeep Barad for the adaptation of HRNet to the Envisat scenario.

References

[1] A. Tatsch, N. Fitz-Coy, S. Gladun, On-orbit servicing: A brief survey, in: Proceedings of the 2006 Performance Metrics for Intelligent Systems Workshop, 2006, pp. 21–23.
[2] M. Wieser, H. Richard, G. Hausmann, J.-C. Meyer, S. Jaekel, M. Lavagna, R. Biesbroek, e.Deorbit mission: OHB Debris removal concepts, in: ASTRA 2015-13th Symposium on Advanced Space Technologies in Robotics and Automation, Noordwijk, The Netherlands, 2015.
[3] J. Davis, H. Pernicka, Proximity operations about and identification of non-cooperative resident space objects using stereo imaging, Acta Astronaut. 155 (2019) 418–425.
[4] V. Pesce, M. Lavagna, R. Bevilacqua, Stereovision-based pose and inertia estimation of unknown and uncooperative space objects, Adv. Space Res. 59 (2017) 236–251.
[5] R. Opromolla, G. Fasano, G. Rufino, M. Grassi, Uncooperative pose estimation with a LIDAR-based system, Acta Astronaut. 110 (2015) 287–297.
[6] S. Segal, P. Gurfil, K. Shahid, In-orbit tracking of resident space objects: A comparison of monocular and stereoscopic vision, IEEE Trans. Aerosp. Electron. Syst. 50 (1) (2014) 676–688.
[7] S. Sharma, J. Ventura, S. D'Amico, Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous, J. Spacecr. Rockets 55 (6) (2018) 1–16.
[8] L. Pasqualetto Cassinis, R. Fonod, E. Gill, Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft, Prog. Aerosp. Sci. 110 (2019).
[9] Lepetit, F. Moreno-Noguer, P. Fua, EPnP: an accurate O(n) solution to the PnP problem, Int. J. Comput. Vis. 81 (2009) 155–166.
[10] L. Ferraz, X. Binefa, F. Moreno-Noguer, Very fast solution to the PnP problem with algebraic outlier rejection, in: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 2014, http://dx.doi.org/10.1109/CVPR.2014.71.
[11] A. Ostrowsky, Solution of Equations and Systems of Equations, second ed., Academic Press, New York, 1966, pp. 183–188.
[12] S. Urban, J. Leitloff, S. Hinz, MLPnP: A rel-time maximum likelihood solution to the perspective-n-point problem, in: Proceedings of the British Machine Vision Conference, vol. 3, 2016.
[13] L. Ferraz, X. Binefa, F. Moreno-Noguer, Leveraging feature uncertainty in the PnP problem, in: Proceedings of the British Machine Vision Conference, Nottingham, UK, 2014, http://dx.doi.org/10.5244/C.28.83.
[14] J. Cui, C. Min, X. Bai, J. Cui, An improved pose estimation method based on projection vector with noise error uncertainty, IEEE Photonics J. 11 (2) (2019).
[15] A. Harvard, V. Capuano, E. Shao, S.-J. Chung, Spacecraft pose estimation from monocular images using neural network based keypoints and visibility maps, in: AIAA Scitech 2020 Forum, Orlando, FL, USA, 2020, http://dx.doi.org/10.1109/AERO.2018.8396425.
[16] M. Kisantal, S. Sharma, T. Park, D. Izzo, M. Martens, S. D'Amico, Satellite pose estimation challenge: Dataset, competition design and results, IEEE Trans. Aerosp. Electron. Syst. (2020).
[17] D. Rondao, N. Aouf, Multi-view monocular pose estimation for spacecraft relative navigation, in: 2018 AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, USA, 2018, http://dx.doi.org/10.2514/6.2018-2100.
[18] V. Capuano, S. Alimo, A. Ho, S. Chung, Robust features extraction for on-board monocular-based spacecraft pose acquisition, in: AIAA Scitech 2019 Forum, San Diego, CA, USA, 2019, http://dx.doi.org/10.2514/6.2019-2005.
[19] S. Sharma, C. Beierle, S. D'Amico, Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks, in: IEEE Aerospace Conference, Big Sky, MT, USA, 2018, http://dx.doi.org/10.1109/AERO.2018.8396425.
[20] S. Sharma, S. D'Amico, Pose estimation for non-cooperative spacecraft rendezvous using neural networks, in: 29th AAS/AIAA Space Flight Mechanics Meeting, Ka'anapali, HI, USA, 2019, http://dx.doi.org/10.1109/AERO.2018.8396425.
[21] J. Shi, S. Ulrich, S. Ruel, CubeSat simulation and detection using monocular camera images and convolutional neural networks, in: 2018 AIAA Guidance, Navigation, and Control Conference, Kissimmee, FL, USA, 2018, http://dx.doi.org/10.2514/6.2018-1604.
[22] S. Sonawani, R. Alimo, R. Detry, D. Jeong, A. Hess, H. Ben Amor, Assistive relative pose estimation for on-orbit assembly using convolutional neural networks, in: AIAA Scitech 2020 Forum, Orlando, FL, USA, 2020, http://dx.doi.org/10.1109/AERO.2018.8396425.
[23] G. Pavlakos, X. Zhou, A. Chan, K. Derpanis, K. Daniilidis, 6-DoF object pose from semantic keypoints, in: IEEE International Conference on Robotics and Automation, 2017.
[24] A. Newell, K. Yang, J. Deng, Stacked hourglass networks for human pose estimation, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), Computer Vision - ECCV 2016, vol. 9912, Springer, Cham, 2016, pp. 483–499.
[25] B. Chen, J. Cao, A. Parra, T. Chin, Satellite pose estimation with deep landmark regression and nonlinear pose refinement, in: International Conference on Computer Vision, Seoul, South Korea, 2019.
[26] T. Park, S. Sharma, S. D'Amico, Towards robust learning-based pose estimation of noncooperative spacecraft, in: AAS/AIAA Astrodynamics Specialist Conference, Portland, ME, USA, 2019.
[27] S. D'Amico, M. Benn, J. Jorgensen, Pose estimation of an uncooperative spacecraft from actual space imagery, Int. J. Space Sci. Eng. 2 (2) (2014) 171–189.
[28] K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation, in: 2019 IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019.

[29] L. Pasqualetto Cassinis, R. Fonod, E. Gill, I. Ahrns, J. Gil Fernandez, Comparative assessment of image processing algorithms for the pose estimation of an uncooperative spacecraft, in: International Workshop on Satellite Constellations & Formation Flying, Glasgow, UK, 2019.
[30] L. Pasqualetto Cassinis, R. Fonod, E. Gill, I. Ahrns, J. Gil Fernandez, CNN-based pose estimation system for close-proximity operations around uncooperative spacecraft, in: AIAA Scitech 2019 Forum, Orlando, FL, USA, 2020, http://dx.doi.org/10.2514/6.2020-1457.
[31] S. Sharma, S. D'Amico, Reduced-dynamics pose estimation for non-cooperative spacecraft rendezvous using monocular vision, in: 38th AAS Guidance and Control Conference, Breckenridge, CO, USA, 2017.
[32] H. Curtis, Orbital Mechanics for Engineering Students, Elsevier, 2005.
[33] M.A. Fischer, R. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Commun. ACM 24 (6) (1981) 381–395.
[34] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.
[35] D. Kingma, J. Ba, Adam: A method for stochastic optimization, in: 3rd International Conference for Learning Representations, San Diego, CA, USA, 2015, http://dx.doi.org/10.2514/6.2018-2100.
[36] B. Tweddle, A. Saenz-Otero, Relative computer vision based navigation for small inspection spacecraft, J. Guid. Control Dyn. 38 (5) (2015) 969–978.
[37] E. Lefferts, F. Markley, M. Shuster, Kalman filtering for spacecraft attitude estimation, J. Guid. Control Dyn. 5 (5) (1982) 417–429.
[38] K. Barad, Robust Navigation Framework for Proximity Operations around Uncooperative Spacecraft (M.Sc. thesis), Delft University of Technology, Delft, The Netherlands, 2020.
[39] J. Solà, Quaternion kinematics for the error-state Kalman filter, 2017, arXiv e-prints arXiv:1711.02508.

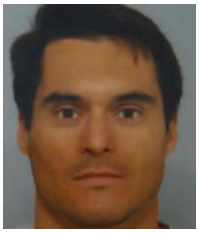


Lorenzo Pasqualetto Cassinis is a Ph.D. candidate in the Space System Engineering section of TU Delft and currently a researcher in the GNC section of ESA's European Space Research and Technology Centre. He graduated with a M.Sc. in Aerospace Engineering in 2017 from TU Delft, and prior to that with a B.Sc. in Aerospace Engineering from the University of Padua in 2015. During his M.Sc., he worked as systems engineer on flight correlation of the DIDO-II CubeSat From December 2017 until May 2018, he also worked at GMV on the validation of the GNC software for the PROBA-3 Mission. His current Ph.D. research relates to monocular-based navigation in debris removal scenarios, with special focus on CNN-based systems. This research is a collaboration between TU Delft, ESA, and Airbus Defence and Space.

Dr. Robert Fonod received the B.Sc. and M.Sc. degrees in Cybernetics from the Technical University of Košice, Slovakia, in 2009 and 2011, respectively, and the Ph.D. degree in Automatic Control from the University of Bordeaux, France, in 2014. He is currently a Research Scientist with the Department of Guidance, Navigation, and Control at the French–German Research Institute of Saint-Louis. He was an Assistant Professor with the Department of Space Engineering at the Delft University of Technology, the Netherlands, and a Postdoctoral Research Fellow with the Department of Aerospace Engineering at the Technion - Israel Institute of Technology, Israel. His research interests are in the area of guidance and estimation of aerospace vehicles, bearings-only target tracking, and model-based fault diagnosis. Dr. Fonod is an Associated Editor for the IEEE Transactions on Aerospace and Electronic Systems.

Dr. Eberhard Gill received a diploma in physics and holds a Ph.D. in theoretical astrophysics from the Eberhard-Karls-University of Tuebingen, Germany. He holds a M.Sc. of Space Systems Engineering from the Delft University of Technology. He has been working as researcher at the German Aerospace Center (DLR) from 1989 to 2006 in the field of precise satellite orbit determination, autonomous navigation and formation flying. He has developed a GPS-based onboard navigation system for the BIRD microsatellite. Dr. Gill has been Co-Investigator on several international missions, including Mars94-96, Mars-Express, Rosetta, Equator-S and Champ, and acted as Principal Investigator on the PRISMA mission. Since 2007, he holds the Chair of Space Systems Engineering at the Faculty of Aerospace Engineering of the Delft University of Technology. In 2013, he has been appointed also as department head Space Engineering at the faculty. Since 2015, he is founding Director of the TU Delft Space Institute.

Dr. Ingo Ahrns received his diploma in computer science in 1996 from the University of Kiel, Germany, and his doctoral thesis in 2000 from the Daimler research center in Ulm, Germany, on the topic of biologically-inspired robot vision. In 2000, he joined the space-robotics department of Airbus Defence and Space. Among many research and development projects, he worked on EUROBOT, DEOS, and the European Lunar Lander, mainly on camera- and LIDAR-based navigation and robot vision activities. In 2018, he started to investigate the use of Deep Learning in space-robotics activities, with focus on CNN based pose-estimation. Since 2013, he became a robot vision expert at Airbus Defence and Space. He currently leads the Autonomous Systems team at Airbus Defence and Space in Bremen, Germany, and coordinates the Airbus Defence and Space robotics R&T cluster.



Dr. Jesús Gil-Fernández received a M.Sc. in Aerospace Engineering from Technical University of Madrid (UPM), another M.Sc. in Theoretical Physics from Universidad Autonoma de Madrid, and a Ph.D. in Aerospace Engineering from UPM. He is GNC engineer at the Guidance, Navigation and Control section in the European Space Agency (ESA) at ESTEC (European Space and Technology Research Center). Prior to ESA he was GNC engineer in GMV for 16 years, the last 4 years he was Head of Interplanetary and NEO missions section.