75th International Astronautical Congress (IAC), Milan, Italy, 14-18 October 2024.

IAC-24-C1.5.1-x89491

Robustness analysis of data driven image processing methods for autonomous navigation with application to the Hera mission

Aurelio Kaluthantrige[1, a], **Mattia Pugliatti**[2, b], **Jinglang Feng**[1, c], **Jesús Gil-Fernández**[3, d], **Francesco Topputo**[4, e]

[1] *Department of Mechanical and Aerospace Engineering, University of Strathclyde, 75 Montrose Street, Glasgow G1 1XJ, United Kingdom.*
[2] *Colorado Center for Astrodynamics Research, University of Colorado Boulder, 3775 Discovery Drive, Boulder, CO 80303, United States.*
[3] *ESA/ESTEC, Keplerlaan 1, Noordwijk, 2201 AZ, The Netherlands.*
[4] *Department of Aerospace Science and Technology, Politecnico di Milano, Via Giuseppe La Masa 34, 20156, Milano, Italy.*
[a] *mewantha.kaluthantrige-don@strath.ac.uk.*
[b] *mattia.pugliatti@colorado.edu.*
[c] *jinglang.feng@strath.ac.uk.*
[d] *jesus.gil.fernandez@esa.int.*
[e] *francesco.topputo@polimi.it.*

Abstract

This work presents a series of functional tests of two data-driven image processing algorithms based on two different convolutional neural networks architectures and designed for the application to the European Space Agency's Hera mission with the target of binary asteroid system (65803) Didymos. The two data-driven methods estimate the position of the centroid of Didymos and its range from the spaccecraft. Through different image datasets and comparative analyses, this work evaluates the two algorithms' performance under conditions of adverse illumination conditions, different shape of the target asteroid and different noise levels of the images, addressing questions on performance deviations and architectural discrepancies, and fine-tuning requirements upon encountering real-world scenarios. The analyses indicate that algorithms with more sophisticated and complex architectures exhibit greater robustness across various contingencies, despite being less accurate in their estimations. Furthermore, the results show that fine-tuning datasets improve the performances of the algorithms in the specific mission scenario they are generated, while reducing the performances in other circumstances.

Keywords: Hera, Image Processing, Guidance Navigation and Control, Autonomous visual based Navigation, Functional Test, Data-driven algorithms

1 Introduction

Space rendezvous operations require high levels of accuracy in terms of relative navigation. When the target is uncooperative, with a highly uncertain dynamical environment and with a large distance from the Earth, spacecraft are equipped with an Autonomous Visual Based Navigation (AVBN) system, which provides a precise visualization of the mission scenario and a navigation system with a high dynamic response to the unknown settings. As such, the design, implementation and testing of an AVBN system is a crucial step in the development of a space rendezvous mission. The AVBN system is integrated with the Guidance, Navigation and Control (GNC) algorithm implemented on the On-Board Computer (OBC) of the spacecraft. The end goal is to have the spacecraft able to estimate autonomously its relative position and/or attitude with respect to the target, and perform manoeuvres (ground-based or calculated on-board) to operate safely around the target [1]. An AVBN system of a space rendezvous mission usually consists of the following three main components [2, 3]:

- An optical sensor that may represent the payload of the spacecraft and that acquires images of the target

body.

- An Image Processing (IP) algorithm that analyses the acquired image and measures pre-defined quantities that gives information on the mission scenario. These may include the range from the target, the relative orientation of the spacecraft with respect to the stellar background, the position of relevant features on the target's surface, etc.;
- A navigation filter that combines the outputs of the IP algorithm with the dynamical environment and provides the best estimate of the relative state of the spacecraft with respect to the target.

In this work we focus on the IP algorithm of an AVBN system. The Design, Development, Validation and Verification (DDVV) strategy for the IP algorithm is incremental and it is based on a chain of different tests that aim to analyze the robustness of the algorithm against potential contingencies [4, 5]. The main tests that are part of a standard DDVV strategy are:

1. **Functional Tests (FT)**: the IP algorithm is tested with synthetic images generated with rendering engines that represent the mission scenario; these images include information on the target, illumination conditions considering the Sun-spacecraft-target relative geometry, relative pose of the spacecraft with respect to the target, and other additional parameters (background noise, distortions etc...). The aim of these tests is to prove that the IP algorithm can provide the required measurements with a level of accuracy high enough for the navigation filter to estimate the spacecraft's state. Additional objectives of these tests are the robustness over external disturbances and noise/distortions of the image. For this purpose, rendering software such as the Planet and Asteroid Natural Scene Generation Utility (PANGU), SurRender, and Blender-based software such as CORTO and SISPO can be used [6, 7, 8, 9, 10].
2. **Model-In-The-Loop (MIL) tests**: the IP algorithm is integrated to the Functional Engineering Simulator (FES) of the GNC prototype of the spacecraft. The FES is a SW environment that includes reference models of the selected GNC solutions and algorithms defined specifically for the mission, and it allows to test the validity of the designed GNC at SW level. The aim of the MIL is to test the interfaces of the IP algorithm with the rest of the GNC models, and more in general that the GNC algorithm is robust to the measurements provided by the IP.
3. **Software-In-The-Loop (SIL) tests**: the GNC algorithm with the embedded IP algorithm is exported to the final programming language that will be used on the OBC of the spacecraft. The aim of the SIL is to verify the correctness of the SW implementation with respect to the FES and to test the interfaces with all the other on-board SW.
4. **Processor-In-The-Loop (PIL) tests**: the validated SW from the previous step is implemented on the qualified processor of the spacecraft. These tests are aimed to profile the different algorithms of the SW and to check the performances in terms of computational time and on-board memory requirements. This test allows to tackle the issue of the HW implementation.
5. **Hardware-In-The-Loop (HIL) tests**: in these tests a representative GNC sensor/actuator is included in the loop, for an AVBN usually being the Functional Model (FUMO) of the camera designed for the mission. The aim of the HIL is to test the robustness of the embedded IP algorithm to the noise, errors and any other electro-optical effect introduced by the camera. This test is usually executed using optical testbenches projecting the scenery on high-resolution screens or in a robotic facility using 3D-printed mock-up of the target body [11, 12, 13, 14].

This work presents a series of FTs of two data-driven-based IP algorithms developed by the authors and presented in [15, 16]. The aim of this work is to validate the functionality and test the robustness of the developed algorithms to multiple mission scenarios and contingencies that are typical of a space rendezvous mission.

In particular, the case study of this work is the European Space Agency (ESA)'s Hera mission during the proximity operations around the target binary asteroid system (65803) Didymos, which consists in a primary body, Didymos, and a secondary body, its moon Dimorphos [17]. During the first phases of the proximity operations the navigation is an AVBN system that relies on an IP algorithm that estimates the position of the Center of Mass (COM) of the primary body from the images captured by the on-board camera. In order to estimate the relative position of the spacecraft with respect to the target, the COM measurements are combined with range measurements from the primary retrieved by the on-board LIDAR (Light Detection and Ranging) system [18, 19].

The application of data-driven methods to the IP algorithm in asteroid rendezvous missions is driven by the possibility to overcome challenges represented by adverse illumination conditions, irregular shape of the target, background noise or the presence of other bodies such as stars or moons that are usually affecting the performances of standard IP algorithms. Recent trends in data-driven based IP algorithms are exploring the use of Artificial Intelligence (AI), particularly deep learning, to address these challenges and bypass traditional techniques. Their utilization in space IP and navigation is becoming more and more relevant given their capability of learning complex features from the provided data [20, 21, 22, 23, 16]. The two data-driven methods analyzed in this work are based on Convolutional Neural Networks (CNNs) with different architectures, number of parameters and with a different set of outputs.

Nevertheless, AI-based methods are not yet in general validated for critical functions of space missions. This is mainly due to the lack of complete mathematical methods to cope with the "reality gap" that separates simulated data with real data. Thus, it remains uncertain how unforeseen changes during the data acquisition process will impact the outcome of AI-based methods. For instance, input data could be blurred, suffer from under/overexposure or contain some noise sources. Therefore, it is important to assess the robustness of AI-based methods against input data alterations [24]. Most of the robustness analysis of AI techniques have been applied considering as inputs adversarial examples [24, 25] and quality distorted data [26, 27]. With the FTs presented in this work, we analyze the performance of the two data-driven methods applied to the IP algorithm of the Hera mission when estimating the position of the COM of Didymos and the range from it, using images representing conditions never encountered during training. In particular, the FTs aim to provide answers to the following research questions: 1) What degree of deviations from design conditions can CNN approaches endure in terms of performance drop? 2) Are there fundamental discrepancies in terms of performance between different architectures? 3) Once Hera arrives at the Didymos system, how many images would be required to fine-tune the data-driven IP algorithms to ensure optimal performances? The first question is assessed by developing a series of challenging datasets that assess network performance with unseen conditions such as high Sun phase angles (Sun-asteroid-spacecraft), different noise levels in the images, and different shape models of the target asteroid system. The second is addressed by performing the analysis with two different convolutional approaches. The third is assessed by performing a series of fine-tuning episodes with incremental dataset sizes. This emulates a possible operational scenario applicable to any data-driven method applied to a space mission, which would require a small subset of images from the real system to be fine-tuned.

While the Hera mission GNC algorithm implemented on the OBC of the spacecraft is going to use the Maximum Correlation with a Lambertian Sphere (MCLS) IP algorithm, the FTs carried in this work represent a fundamental validation step toward the applicability of data-driven methods in the AVBN of asteroid rendezvous missions [5].

This paper is structured as follows. Section 2 describes in detail the data-driven methods, the case study and the FTs carried in this work. Section 3 shows the obtained results and discusses the applicability of the developed methods to the Hera mission. Finally, Section 4 concludes this work.

2 Methodology

This section describes the methodology applied in this work to validate through FTs the data-driven methods developed by the authors. Firstly, the case study is presented, in Section 2.1. Then the data-driven methods are described in Section 2.2. The first data-driven method relies on a CNN-based architecture called High Resolution Network (HRNet), developed in [15] and it is briefly outlined in Section 2.2.1. The second data-driven method relies on an architecture applied to the AVBN system of Milani, one of the CubeSats of the Hera mission [28, 29]. It is developed in [16] and it is introduced in Section 2.2.2 . We refer to the first data-driven method as M1 and to the second as M2. Finally, the FTs with their objectives and the datasets used are detailed in Section 2.4. To facilitate the discussion, we refer to Didymos as B1 and to Dimorphos as B2. The geometrical center of B1 is considered its COM for simplicity.

2.1 Case Study

The proximity operations of the Hera mission around the target binary asteroid system Didymos represent the case study of this work. Hera serves as Europe's contribution to the Asteroid Impact and Deflection Assessment (AIDA) international collaboration with NASA. AIDA

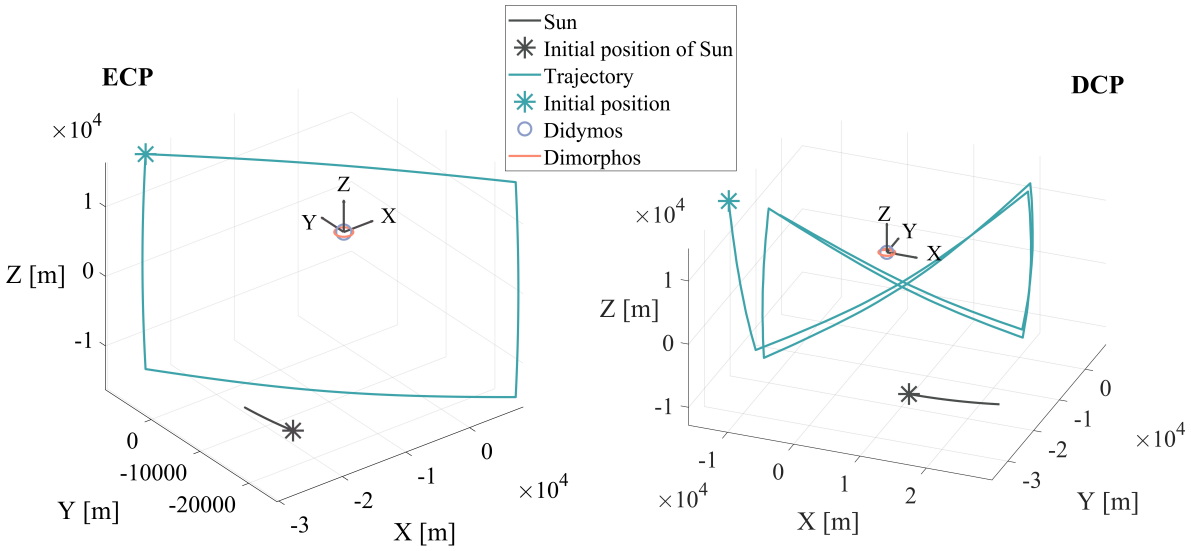

Figure 1: ECP and DCP trajectories.

aims to demonstrate asteroid deflection using NASA's kinetic impactor Double Asteroid Redirection Test (DART) spacecraft, which achieved successfully its objective on September 26, 2022. Hera will rendezvous with the target asteroid in early 2027 to study its physical and dynamic properties, including the impact crater and the momentum transfer efficiency [30, 31, 29]. The current knowledge on the shapes of B1 and B2 is provided by DART's latest observations, and it is that they are two oblate ellipsoids, with the extent along the principal axes and the respective uncertainties given in Table 1 [32, 33]. The table shows the data retrieved by DART prior to the impact, as the shape resulted from the collision will be only resolved once the Hera spacecraft will reach the asteroid [34].

The properties of the Hera on-board Asteroid Framing Camera (AFC) shown in Table 2 are used for the generation of the synthetic images [35, 36].

The images of the FTs are generated with the aim to prove the robustness of the algorithm in different scenarios, including two specific phases of the proximity operations, the Early Characterization Phase (ECP) and the subsequent Detailed Characterization Phase (DCP), shown in Fig. 1 (courtesy of ESA). The two trajectories are represented in the Target Body reference frame (TB), which uses B1's COM as origin, the x-axis parallel to the one of the Earth-centered inertial coordinate frame and the xy-plane coplanar to the orbit of B2. During the ECP, the navigation and the attitude profile is ground based and images of the target are captured and downlinked to Earth in order to tune the on-board IP algorithm to optimize its performances. The range of the spacecraft from B1 varies from $\sim$ 20 km to $\sim$ 30 km and the Sun phase angle ranges from $\sim 45°$ to $\sim 80°$. During the DCP, the navigation is fully autonomous and it relies on the tuned IP algorithm. The range of the spacecraft from B1 varies from $\sim$ 9 km to $\sim$ 20 km and the Sun phase angle ranges from $\sim 0°$ to $\sim 80°$ [37, 38]. The orbit of Dimorphos prior to DART's impact is considered as the orbital changes resulted in the collision are not affecting the FTs carried out in this work.

2.2 Data-driven methods

2.2.1 M1

M1's pipeline is shown in Fig. 2 and it consists of the following blocks:

1. *Pre-Processing*: The input images are scaled down (256×256 px) and the magnitude of each pixel is normalized for the convergence of the HRNet;
2. *HRNet*: The HRNet block applies the HRNet to the input image in order to regress the position of the COM of B1, the COM of B2, and 24 points on the illuminated border of B1. The latter are used together with the COM of B1 to derive its range from the spacecraft geometrically, by approximating the shape of B1 to a sphere of diameter $D = 780$ m. The output of the HRNet block is a set of 26 64×64

Parameter	B1	B2
Extent along principal axis x $[m]$	849 ± 5.6	177 ± 1.2
Extent along principal axis y $[m]$	851 ± 5.6	174 ± 1.2
Extent along principal axis z $[m]$	620 ± 5.6	116 ± 1.2

Table 1: Shapes of B1 and B2 prior to DART's impact [32, 33].

Field Of View (FOV)	Focal Length	Aperture	Image size	Pixel Size
$5.5°$	$10.6\ cm$	$2.5\ cm$	$1024 \times 1024\ px$	$10\ \mu m$

Table 2: AFC's properties [35, 36].

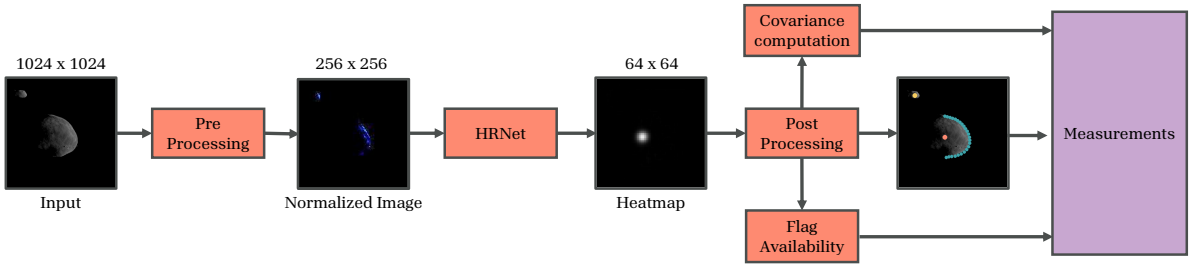

Figure 2: M1 data-driven method.

px heatmaps, each one associated with the regressed point [39, 40]. A heatmap manifests as a cluster of white pixels encircling the predicted point, indicating the degree of accuracy in determining its position. The denser and more concentrated the heatmap, the more precise the estimation of the point's location.

3. *Post-Processing*: Each heatmap is denoised and analyzed to calculate the exact position of the desired point. A statistical population of pixels of the heatmap around the desired point is extracted.

4. *Flag Availability*: If the pixels of a heatmap are less intense than a pre-defined threshold, the associated point is not given as an output and the corresponding measurement is not available.

5. *Covariance Computation*: If the pixels of the heatmap are more intense than the pre-defined threshold, the associated point is given as an output, the measurement is available and its covariance is derived by the magnitude and shape of the heatmap.

Further details about this algorithm can be found in [41, 15].

2.2.2 M2

M2's pipeline is shown in Fig. 3 while the CNN architecture is illustrated in Table 3 using TensorFlow 2.10 notation. The architecture's hyper-parameters are a result of a thorough search based on the use of extreme-learning machine methods and hierarchical grid search, an approach that can be found in [42].

The convolutional portion of the architecture is divided into 5 depth levels that make extensive use of dilated convolutions of 1, 2, and 3 rates (represented at each $i-th$ depth level by the C_{i1}, C_{i2}, and C_{i3} layers), concatenation, activation, and max-pooling layers. The head of the architecture is represented by a single dense output layer connected directly to the fully connected $FC5$ layer. Dropout is applied on this layer with a probability of 15%.

Differently from M1, M2 is an end-to-end architecture, since it generates an output vector consisting of B1 COM components, range from B1, and Sun phase angle. Moreover, M2 uses as input a cropped and/or resized version of

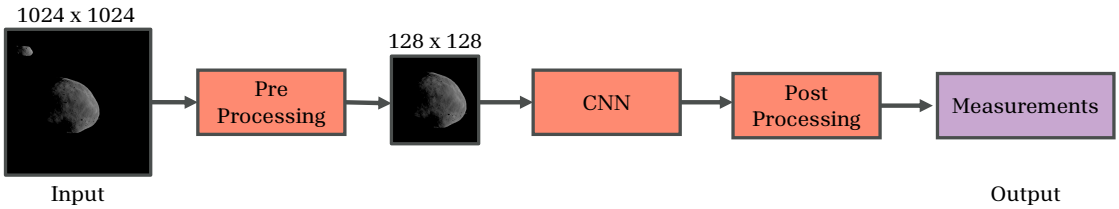

Figure 3: M2 data-driven method.

the original 1024×1024 px image. Each image acquired by the camera, thus needs to be pre-processed outside the network before being considered as an input. During training, labels need to be modified accordingly to reflect the reduced input size.

While M1 uses resized 256×256 px images from the original 1024×1024 px, M2 needs a more elaborate image-label preprocessing to reduce each input to 128×128 px images. This is performed with an adaptable cropping algorithm that first produces a region of interest on the image which is one of four possible sizes (128×128 px, 256×256 px, 512×512 px, and 1024×1024 px depending on how the body appears in the image at a specific range) and then generates a resized 128×128 px image. This process also changes the values of the COM of B1 and range, which need to be adjusted both during training and testing. Further details about this algorithm can be found in [16, 42].

2.3 Overview of data-driven methods

Table 4 reports the main parameters and characteristics of the IP algorithms described in this section. It can be seen that M1 has roughly 8 times more parameters than M2, due to its heavier and more complex architecture. This also affects the higher Average Computational Time (ACT) required by M1 to process a single image on a Zynq-7000 System on a Chip (SoC) processor considered in this work as representative of a typical OBC. It can also be noted that M2 is an end-to-end data-driven method and derives the range from B1 directly with the CNN architecture, while M1 takes advantage of B1's shape and derives it geometrically.

2.4 FTs analysis

The FTs run in this work are designed to test the robustness and performances of the developed IP algorithm against multiple adverse conditions that could occur during mission operations. Specifically, the objective of the FTs is to analyze the capability of the trained IP algorithm to generalize their solution when facing conditions never seen during training. In particular, the conditions analyzed in this work are of different illumination, different shape of the target body, presence/absence of B2, different Signal-to-Noise ratio, and, more in general, different images than the ones generated synthetically and used on ground. Furthermore, we analyze the applicability of the algorithm during Hera's proximity operations. Since the baseline Hera AVBN system consists of the measurements of the position of the COM of B1 given by the on-board IP MCLS algorithm and the range from B1 given by the on-board LIDAR system, only these two outputs of the developed IP algorithms are analyzed with the FTs. The next section describes in detail the different types of datasets generated in this work.

2.4.1 Training and Testing Datasets

In this work, several datasets are generated to train and test the data-driven methods. A summary of their main properties is represented in Table 6 and Table 7, while a representation of the distribution of the dataset in the space surrounding the Didymos system is illustrated in Fig. 4. Table 5 reports the values of the length of the principal axes of the different shape models of B1 and B2 used in this work. S_1 represents the current knowledge of these shape models without uncertainties as shown in Table 1 [32]. The shape models of S_2 instead, considers as

Layer	**Type**	**Output Shape**	**Parameters**	**Connected to**
I	InputLayer	$B, 128, 128, 1$	0	-
C_{11}	Conv2D	$B, 128, 128, 16$	160	I
C_{12}	Conv2D	$B, 128, 128, 16$	160	I
C_{13}	Conv2D	$B, 128, 128, 16$	160	I
$CC1$	Concatenate	$B, 128, 128, 48$	0	C_{11}, C_{12}, C_{13}
$A1$	Activation	$B, 128, 128, 48$	0	$CC1$
$P1$	MaxPooling2D	$B, 64, 64, 48$	0	$A1$
C_{21}	Conv2D	$B, 64, 64, 32$	13856	$P1$
C_{22}	Conv2D	$B, 64, 64, 32$	13856	$P1$
C_{23}	Conv2D	$B, 64, 64, 32$	13856	$P1$
$CC2$	Concatenate	$B, 64, 64, 96$	0	C_{21}, C_{22}, C_{23}
$A2$	Activation	$B, 64, 64, 96$	0	$CC2$
$P2$	MaxPooling2D	$B, 32, 32, 96$	0	$A2$
C_{31}	Conv2D	$B, 32, 32, 64$	55360	$P2$
C_{32}	Conv2D	$B, 32, 32, 64$	55360	$P2$
C_{33}	Conv2D	$B, 32, 32, 64$	55360	$P2$
$CC3$	Concatenate	$B, 32, 32, 192$	0	C_{31}, C_{32}, C_{33}
$A3$	Activation	$B, 32, 32, 192$	0	$CC3$
$P3$	MaxPooling2D	$B, 16, 16, 192$	0	$A3$
C_{41}	Conv2D	$B, 16, 16, 128$	221312	$P3$
C_{42}	Conv2D	$B, 16, 16, 128$	221312	$P3$
C_{43}	Conv2D	$B, 16, 16, 128$	221312	$P3$
$CC4$	Concatenate	$B, 16, 16, 384$	0	C_{41}, C_{42}, C_{43}
$A4$	Activation	$B, 16, 16, 384$	0	$CC4$
$P4$	MaxPooling2D	$B, 8, 8, 384$	0	$A4$
C_{51}	Conv2D	$B, 8, 8, 256$	884992	$P4$
C_{52}	Conv2D	$B, 8, 8, 256$	884992	$P4$
C_{53}	Conv2D	$B, 8, 8, 256$	884992	$P4$
$CC5$	Concatenate	$B, 8, 8, 768$	0	C_{51}, C_{52}, C_{53}
$A5$	Activation	$B, 8, 8, 768$	0	$CC5$
$P5$	MaxPooling2D	$B, 4, 4, 768$	0	$A5$
$FC5$	Flatten	$B, 12288$	0	$P5$
DO	Dropout	$B, 12288$	0	$FC5$
D	Dense	$B, 3$	49156	DO

Table 3: M2 Architecture, made of 3563907 parameters, all of which are trainable.

Parameter	**M1**	**M2**
Number of model parameters	28.5 M	3.6 M
Weight	109 MB	13.6 MB
ACT on a Zynq-7000 SoC	165 s	9.94 s
Outputs	COM B1, COM B2, Range from B1, Associated covariances, Flag if measurements are available or not	COM B1, Range from B1, Sun phase angle
Range estimation	Derived geometrically	Estimated from network

Table 4: Overview of the architectures of M1 and M2.

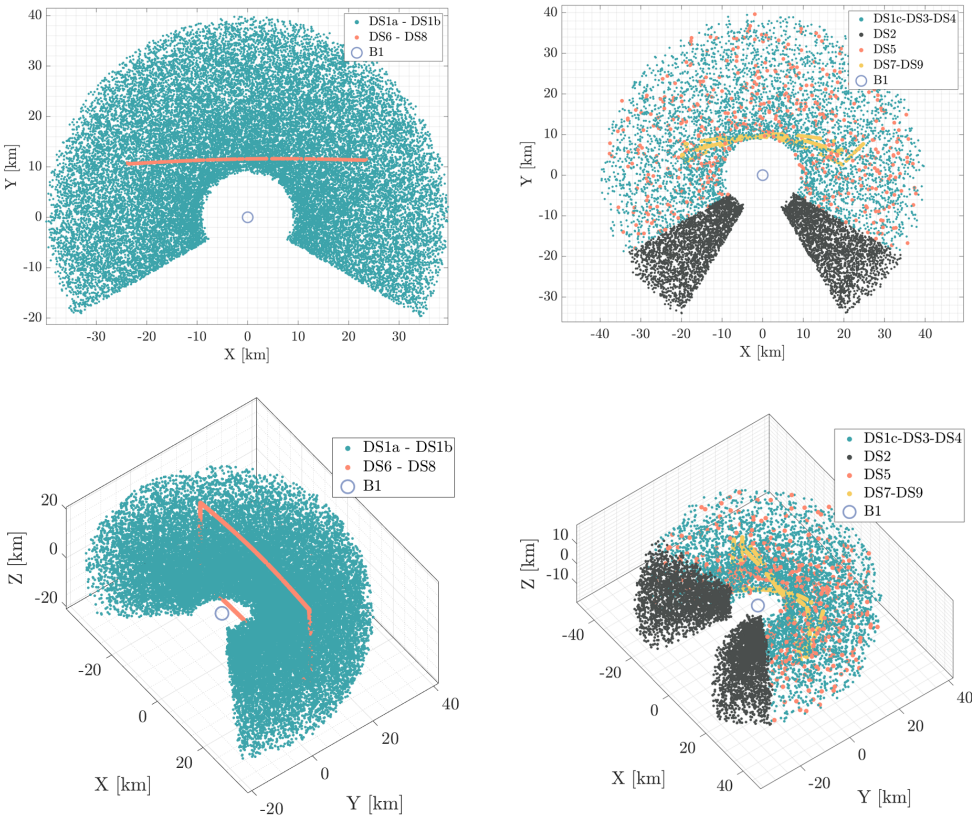

Figure 4: Visualization of the training (left column), and testing (right column) datasets in TB. The Sun is illuminating the asteroid system from the +Y axis.

principal axes the ones retrieved from radar observations, prior to DART's arrival when B1's shape was thought to be more spherical [33]. S_3 represents an expanded version of S_1 in all axes. While the first three shape models are generated with the software Blender, the last one, S_4, is made with the software PANGU and it uses the same principal axes length of S_1.

The same pipeline is used to generate all the images. However, all but $DS6$ and $DS7$ are generated in CORTO [9], while PANGU is instead used to generate $DS8$ and $DS9$. CORTO stands for Celestial Object Rendering TOol and it is an open-access tool that uses Blender to generate high-fidelity, large, annotated datasets of celestial bodies. The tool represents a versatile and comprehen-

Shape models	$B1_x$ $[m]$	$B1_y$ $[m]$	$B1_z$ $[m]$	$B2_x$ $[m]$	$B2_y$ $[m]$	$B2_z$ $[m]$	Modeling software
S_1	849	851	620	177	174	116	Blender
S_2	821	823	786	202	159	134	Blender
S_3	861	862	626	178	175	117	Blender
S_4	849	851	620	177	174	116	PANGU

Table 5: Axes elongation of the shape models used in this work.

sive solution for generating synthetic images of celestial bodies, aiding the development and validation of image processing and navigation algorithms for space missions.

The training datasets are represented in Table 6 while the testing datasets are in Table 7. $DS1_a$ and $DS1_b$ are used to train M2 and M1 respectively. The trained M1 and M2 are then tested with $DS1_c$, $DS2$, $DS3$, $DS4$ and $DS5$. The fine-tuning of M1 and M2 is carried with $DS6$ and $DS8$ and the respective tests are done with $DS7$ and $DS9$ subsequently.

The geometric distributions of the locations where the images are generated are illustrated in Fig. 4. The distribution of $DS1_a$, $DS1_b$, $DS1_c$, $DS3$, $DS4$, and $DS5$ is characterized by random points between $10\ km$ and $40\ km$ from Didymos, with Sun phase angles ranging from 0° to 120°, and absolute values of elevation angle with respect to Didymos' equator between 0° and 30°. The training dataset $DS1_b$ differs from $DS1_a$ simply by the absence of B2 from the images, which is necessary to train M1. $DS3$ and $DS4$ are used to assess the IP performance with a different B1 shape model with principal axes shown in Table 5. While $DS3$ represents only B1 and is used to isolate the error introduced by the different shape of the target, $DS4$ represents both B1 and B2, therefore including also the dependence of the IP algorithms performances on the presence of B2.

$DS5$ represents a small subset of the first 500 points of $DS1$ in which the settings of the artificial noise are varied. For the noise model used, the reader is directed to [9]. In this work, the number of pixels considered in the horizontal motion blur (ν_{mb}), a generic isotropic blur (ν_b), a gamma correction factor (γ), mean (ν_μ) and variance (ν_σ) of Gaussian noise are sampled with random uniform distributions according to the extremal values reported in Table 8 (note, however, that ν_μ and ν_σ are sampled in logarithmic scale).

The subsets of $DS5$ from 1 to 4 are referred to as "high blur" (B) while those from 5 to 9 are considered "low blur" (b). The subsets 1, 2, 5, and 6 are referred as "low gamma" (g), the subsets 3, 4, 7, and 8 are referred as "high gamma" (G), while the subset 9 is considered "nominal gamma"(g0). Similar distinctions are also made for "low mean and low variance" (l), such as for subsets 2, 3, 5, 7, and 9, and "high mean and high variance" (L), such as for subsets 1, 4, 6, and 8. Lastly, the distribution of $DS2$ points differs from all those described above only for one condition: the illumination conditions are adverse, with the Sun phase angles ranging from 120° to 150°.

Finally, $DS6$ and $DS8$ distributions represent points scattered uniformly across the ECP trajectory of the Hera mission. The same is done for $DS7$ and $DS9$ with the DCP. However, while $DS6$ and $DS8$ are rendered with CORTO using shape models generated with Blender, $DS7$ and $DS9$ are using shape models with the same length of the principal axes as S_1 but generated with PANGU. Both $DS6$ and $DS8$ are used in eight, separate, fine-tuning episodes in which the number of images used is doubled each time from 8 to 1024. These datasets mimic real mission scenarios in which a limited amount of images could be available to fine-tune a data-driven method. The eight different fine-tuned networks are then tested with $DS7$ and $DS9$ respectively, as they would be deployed in the next phase of the Hera mission, to assess the impact of the fine-tuning performed during the ECP.

All datasets except for $DS9$ consider a random relative attitude of the spacecraft with respect to B1 for each image, showing therefore illumination coming from any possible direction. In $DS9$ the Hera spacecraft's body reference frame is considered, having the Sun-B1 vector always lying on the horizontal axis of the image plane, thus showing the target asteroid illuminated from the right side.

Acronym	Number of images	Description	Shape model	Image generation tool
$DS1_a$	30000	Nominal scenario	S_1	CORTO
$DS1_b$	30000	Nominal scenario, without B2	S_1	CORTO
$DS6$	$8, 16, 32, 64, 128, 256, 512, 1024$	ECP, new shape	S_3	CORTO
$DS8$	$8, 16, 32, 64, 128, 256, 512, 1024$	ECP, new shape	S_4	PANGU

Table 6: Summary of the properties of the training datasets.

Acronym	Number of images	Description	Shape model	Image generation tool
$DS1_c$	5000	Nominal scenario	S_1	CORTO
$DS2$	5000	High Sun phase angles	S_1	CORTO
$DS3$	5000	Nominal scenario, new shape, without B2	S_2	CORTO
$DS4$	5000	Nominal scenario, new shape, with B2	S_2	CORTO
$DS5$	500×9	Nominal scenario, different noises	S_1	CORTO
$DS7$	450	DCP, new shape	S_3	CORTO
$DS9$	450	DCP, new shape	S_4	PANGU

Table 7: Summary of the properties of the testing datasets.

Subset	ν_{mb}	ν_b	γ	ν_μ	ν_σ	Acronym
Nominal values	[0.10, 2.00]	0.5	[0.9,1.1]	[1e-4,1e-1]	1e-4	-
1	[1.00, 2.00]	[1.00, 2.00]	[0.70, 1.00]	[1e-3, 1e-2]	[1e-3, 1e-2]	BgL
2	[1.00, 2.00]	[1.00, 2.00]	[0.70, 1.00]	[1e-5, 1e-3]	[1e-5, 1e-3]	Bgl
3	[1.00, 2.00]	[1.00, 2.00]	[1.00, 1.30]	[1e-5, 1e-3]	[1e-5, 1e-3]	BGl
4	[1.00, 2.00]	[1.00, 2.00]	[1.00, 1.30]	[1e-3, 1e-2]	[1e-3, 1e-2]	BGL
5	[0.20, 1.00]	[0.20, 1.00]	[0.70, 1.00]	[1e-5, 1e-3]	[1e-5, 1e-3]	bgl
6	[0.20, 1.00]	[0.20, 1.00]	[0.70, 1.00]	[1e-3, 1e-2]	[1e-3, 1e-2]	bgL
7	[0.20, 1.00]	[0.20, 1.00]	[1.00, 1.30]	[1e-5, 1e-3]	[1e-5, 1e-3]	bGl
8	[0.20, 1.00]	[0.20, 1.00]	[1.00, 1.30]	[1e-3, 1e-2]	[1e-3, 1e-2]	bGL
9	[0.20, 0.21]	[0.20, 0.21]	[0.99, 1.01]	[1e-6, 1e-5]	[1e-6, 1e-5]	bg0l

Table 8: Summary of the noise properties varied in the different subsets of $DS5$. For comparison, the first row represents the nominal values used in all other datasets.

3 Results

In this section, the results of the FTs of the developed data-driven methods are presented. As mentioned, the performances of the algorithms on the testing datasets are analyzed only for the estimation of the position of the COM of B1 and the range from B1. The metrics defined in Eq. 1 and Eq. 2 are used to evaluate the performances of each data-driven method for each FTs.

$$\varepsilon_{CoM} = CoM^e - CoM^t \tag{1}$$

$$\varepsilon_\rho = \rho^e - \rho^t \tag{2}$$

where CoM and ρ indicate the COM of B1 and the range from B1 respectively, while the superscripts t and e indicate respectively the ground truth and the estimated value.

ε_{CoM} represents the error of the estimated position of the COM in the image plane and is calculated in px while ε_ρ represents the error of the estimated range and is calculated in km. Since the position of the COM of B1 on the image is defined by its (u, v) coordinates on the image plane, ε_{CoM} is a vector with coordinates $(\varepsilon^u_{CoM}, \varepsilon^v_{CoM})$, and its norm is ε^n_{CoM}. Lastly, $\varepsilon^{\%}_\rho$ is used to represent ε_ρ as relative percentage error with respect to the true range ρ^t, as described in Eq. 3

$$\varepsilon^{\%}_\rho = \frac{\rho^e - \rho^t}{\rho^t} \cdot 100\% \tag{3}$$

In order to assess the robustness of the data-driven methods against the various noises, different shape models and adverse illumination conditions, the results of $DS2$, $DS3$, $DS5$, $DS7$ and $DS9$ are compared with the results obtained on the nominal scenario of the images of $DS1_c$. The results of $DS4$ are compared to the ones of $DS3$, to analyze the robustness of the algorithm to the presence of B2.

As mentioned in Section 2.2.1, the availability of the measurements of M1 depends on the intensity of the heatmap generated by the HRNet from an input image. This is not the case for M2, that always converge to a solution. Therefore, it is relevant to assess the convergence of M1 for each dataset, as shown in Table 9, which shows for each dataset the amount of images not solved by M1. For $DS5$, $DS7$ and $DS9$ the best and worst subset results are shown. It can be seen that the worst scenarios are given by the fine-tuning tests $DS7$ and $DS9$ for which up to 48.67% of the images are not solved by M1. In these cases, the heavier architecture of M1 seems to exhibit a higher inertia to the fine-tuning. In order for M1 to be more confident with the provided solution and thus, exhibit intense heatmaps, M1 would require additional images for the fine-tuning.

Dataset	**Absolute** $[-]$	**Rate** $[\%]$
$DS1_c$	195	3.90
$DS2$	751	15.02
$DS3$	359	7.18
$DS4$	350	7.00
$DS5$	7-47	1.40-9.40
$DS7$	31-151	6.89-33.56
$DS9$	37-219	8.22-48.67

Table 9: Images not solved by M1 in absolute and relative terms.

3.1 From $DS1_c$ to $DS4$

Fig. 5, Fig. 6, Fig. 7 and Fig. 8 illustrate the performances of M1 and M2 on the estimation of the position of the COM of B1 and the range from B1 for the testing datasets $DS1_c$, $DS2$, $DS3$ and $DS4$, respectively. Table 10 reports the value of each mean μ and standard deviation σ of the metric in the curly brackets. It can be seen from the values of ε^u_{CoM} and ε^v_{CoM} that both methodologies introduce a small bias in estimating the position of the COM of B1. Independently from the shape model, noise condition, and the direction of the incoming light, M1 underestimates CoM^u, while overestimating CoM^v. On the other hand, M2 overestimates only CoM^v, but with higher values compared to M1.

3.1.1 $DS1_c$: Nominal scenario

It can be seen from the top left plot of Fig. 5 that the error made by M1 on estimating the range from B1 is biased towards negative values ($\mu\{\varepsilon_\rho\} = -2.77\ km$), while the error made by M2 is distributed around 0 ($\mu\{\varepsilon_\rho\} = -0.05\ km$). Moreover, the error of the estimates made by M2 is less dispersed than those of M1, with a standard deviation respectively of $\sigma\{\varepsilon_\rho\} = 1.73\ km$ and $\sigma\{\varepsilon_\rho\} = 1.10\ km$.

The worse performance of M1 on ε_ρ is due to the inaccurate approximation of the shape of B1 to a sphere to derive geometrically the range. This is also visible in the bottom left plot of Fig. 5, where the percent error of ε_ρ is plotted against the ground truth value of the range, showing that the error made by M1 is higher and spread more widely. Nevertheless, the bottom left plot also highlights that the range estimate of both algorithms does not exhibit any relevant trend with respect to the true range.

The top right plot of Fig. 5 shows the distribution of ε^n_{CoM} for M1 and M2. It can be seen that the performances are very similar, with the mean value of ε^n_{CoM} obtained by M1 slightly higher ($\mu\{\varepsilon^n_{CoM}\} = 12.73\ px$ for M1, and $\mu\{\varepsilon^n_{CoM}\} = 10.67\ px$ for M2), with a standard deviation of $\sigma\{\varepsilon^n_{CoM}\} = 7.79\ px$ and $\sigma\{\varepsilon^n_{CoM}\} = 8.66\ px$ respectively. This is also visible in the bottom right plot of Fig. 5, where the distribution of ε_{CoM} is plotted with its (u, v) coordinates in a neighborhood of $64 \times 64\ px$ around the ideal results of $\varepsilon_{CoM} = (0, 0)$.

In conclusion, in the nominal scenario where the testing dataset is similar to the training one, M2 is more accurate

Dataset	**IP**	$\mu(\sigma)\{\varepsilon^u_{CoM}\}[px]$	$\mu(\sigma)\{\varepsilon^v_{CoM}\}[px]$	$\mu(\sigma)\{\varepsilon^n_{CoM}\}[px]$	$\mu(\sigma)\{\varepsilon_\rho\}[km]$	$\mu(\sigma)\{\lvert\varepsilon_\rho\rvert\}[km]$
$DS1_c$	M1	-7.24 (-5.75)	7.82 (8.72)	12.73 (7.79)	-2.77 (1.73)	2.77 (1.72)
$DS1_c$	M2	1.80 (0.44)	9.69 (9.57)	10.67 (8.66)	-0.05 (1.10)	0.79 (0.77)
$DS2$	M1	-7.12 (-7.55)	20.31 (21.39)	25.99 (17.38)	-4.91 (4.29)	4.93 (4.27)
$DS2$	M2	1.98 (2.05)	38.18 (40.06)	44.07 (33.60)	-2.45 (3.38)	2.89 (3.01)
$DS3$	M1	-7.47 (-6.14)	17.02 (17.07)	20.56 (15.86)	1.22 (1.37)	1.41 (1.18)
$DS3$	M2	1.97 (0.48)	20.35 (19.30)	21.07 (18.62)	2.82 (1.51)	2.83 (1.50)
$DS4$	M1	-7.49 (-6.16)	17.18 (17.11)	20.57 (16.08)	1.19 (1.37)	1.39 (1.17)
$DS4$	M2	2.15 (0.62)	19.80 (18.98)	20.70 (18.14)	2.79 (1.51)	2.80 (1.50)

Table 10: Performances of M1 and M2 on the estimation the position of the COM of B1 and the range from B1 for $DS1_c$, $DS2$, $DS3$ and $DS4$.

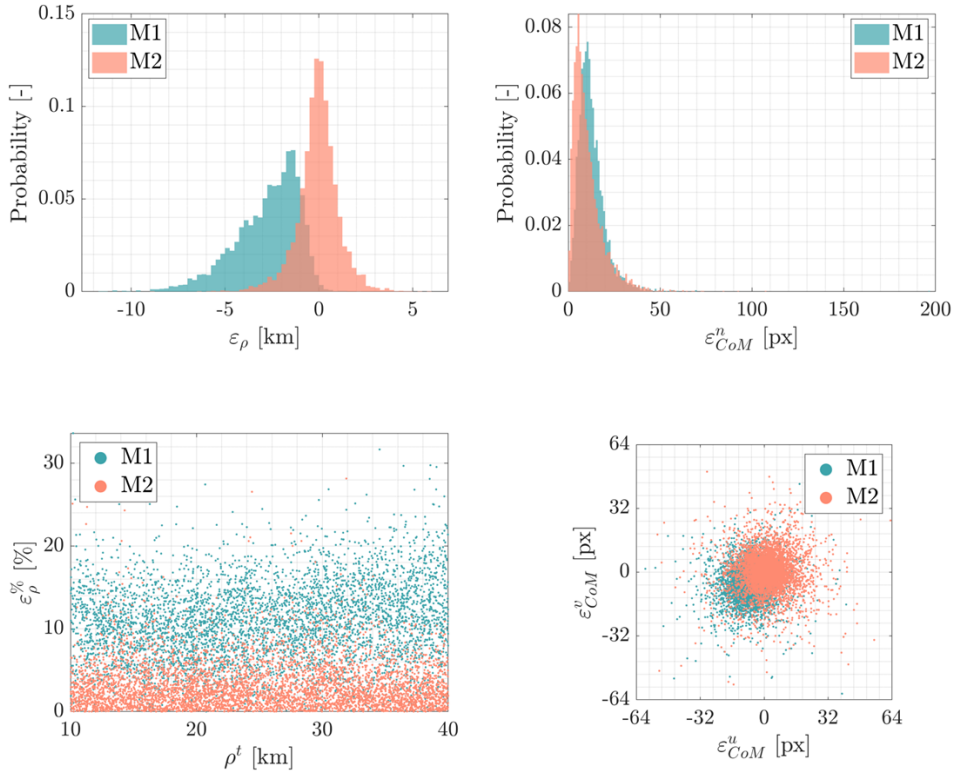

Figure 5: Range (left column) and centroiding (right column) error results of M1 and M2 for $DS1_c$.

and precise than M1 in estimating both range and COM coordinates.

3.1.2 DS2: High Sun phase angles

It can be seen from Table 10 and from Fig. 6 that the adverse illumination conditions represented by the high Sun phase angles of $DS2$ affect the performances of M1 and M2, shown by the higher values of $\mu\{\varepsilon_\rho\}$ and $\mu\{\varepsilon^n_{CoM}\}$. As in $DS1_c$, the error distribution on the range estimation is biased towards negative values for M1 ($\mu\{\varepsilon_\rho\} = -4.91\ km$, $\sigma\{\varepsilon_\rho\} = 4.29\ km$) with respect to M2 ($\mu\{\varepsilon_\rho\} = -2.45\ km$, $\sigma\{\varepsilon_\rho\} = 3.38\ km$). For the estimation of the position of the COM of B1, M1 performs better than M2, with a $\mu\{\varepsilon^n_{CoM}\} = 25.99\ px$ for M1 and $\mu\{\varepsilon^n_{CoM}\} = 44.07\ px$ for M2, and with a standard devia-

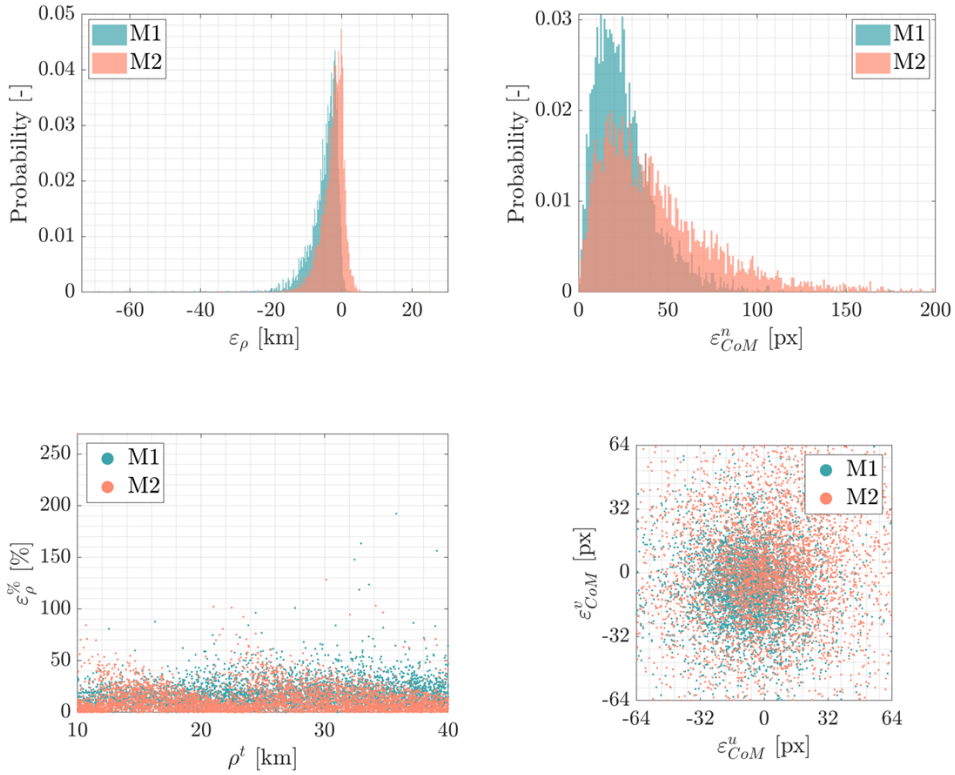

Figure 6: Range (left column) and centroiding (right column) error results of M1 and M2 for $DS2$.

tion of $\sigma\{\varepsilon^n_{CoM}\} = 17.38\ px$ and $\sigma\{\varepsilon^n_{CoM}\} = 33.60\ px$ respectively. The worse performances of the two methods with respect to the nominal scenario can be seen from the bottom left and the bottom right plots of Fig. 6, which show that ε_ρ and ε^n_{CoM} are less precise compared to the nominal scenario.

To conclude, in the scenario in which higher phase angles are considered, both methods suffer a considerable drop in performance compared to the results in $DS1_c$. However, M1 seems to suffer a smaller drop compared to M2, indicating a higher degree of robustness. From the results presented in this section, it appears that the visible edge of the body in such challenging illumination conditions works better with the geometrical derivation strategy performed with the HRNet in M1 than with the end-to-end approach of M2. This is particularly visible looking at the performance on the range, which greatly degrade for M2, while they only mildly degrade for M1 with respect to the ones obtained with $DS1_c$. However, it is also noted that M1 does not converge for a higher number of cases (∼4 times higher than in $DS1_c$, see Table 9).

3.1.3 DS3: Nominal scenario, new shape of B1, without B2

It can be seen from Table 10 and from Fig. 7 that the different shape of B1 affect slightly the performances of M1 and M2. Contrarily to the results obtained in $DS1_c$ and $DS2$, the error distribution on the range estimation is now biased towards positive values for M1 ($\mu\{\varepsilon_\rho\} = 1.22\ km$, $\sigma\{\varepsilon_\rho\} = 1.37\ km$), with a higher accuracy than M2 ($\mu\{\varepsilon_\rho\} = 2.82\ km$, $\sigma\{\varepsilon_\rho\} = 1.51\ km$). Considering that in this dataset the shape of B1 is more spherical, the geometrical derivation of the range from its shape improves the performances of M1, while reducing the ones of M2. For the estimation of the position of the COM of B1, M1 performs slightly better than M2, with a $\mu\{\varepsilon^n_{CoM}\} = 20.56\ px$ for M1 and $\mu\{\varepsilon^n_{CoM}\} = 21.07\ px$ for M2 and with a standard deviation of $\sigma\{\varepsilon^n_{CoM}\} = 15.86\ px$ and $\sigma\{\varepsilon^n_{CoM}\} = 18.62\ px$ respectively. The better performances of M1 in the range estimation can be seen from the bottom left and bottom right plots of Fig. 7, which show that ε_ρ and ε^n_{CoM} have values closer to 0 with respect to M2.

To conclude, the different shape of B1 tested in $DS3$ seems to have a less drastic effect than unforeseen illu-

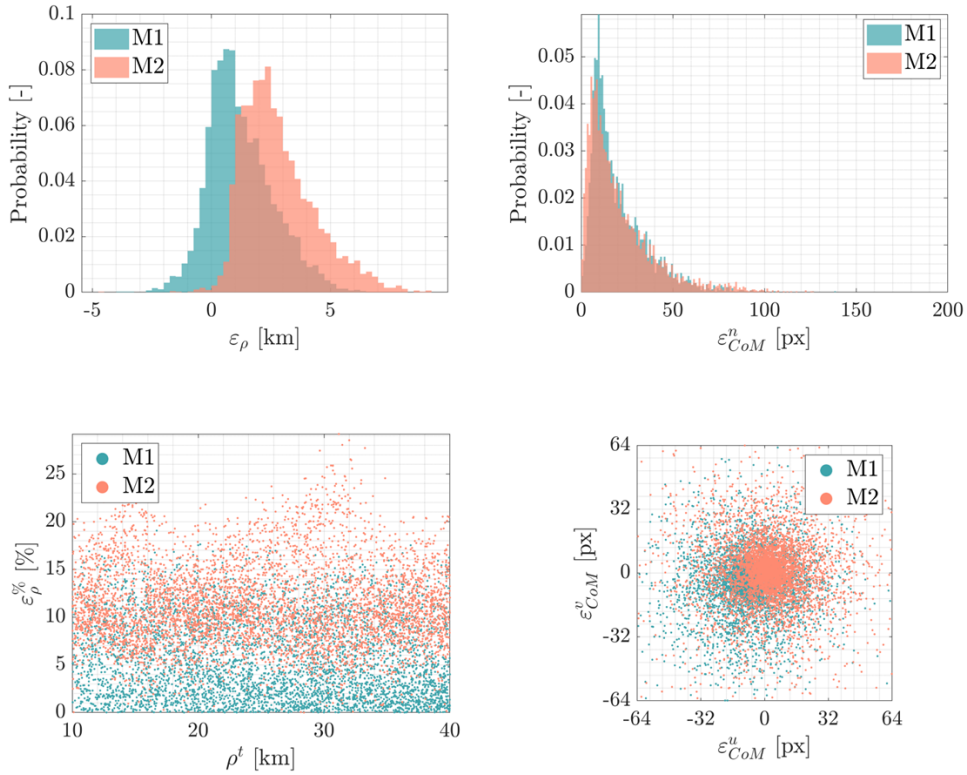

Figure 7: Range (left column) and centroiding (right column) error results of M1 and M2 for $DS3$.

mination conditions tested in $DS2$ as both M1 and M2 performance are closer to those in $DS1_c$. The different shape models seem to favor both the accuracy and precision of the range estimate of M1 compared to M2. This is attributed to the more spherical shape of B1 in this dataset, which performs better in the apparent diameter formulation used by M1 compared to the end-to-end approach used in M2. Centroiding performance on the other hand are basically very similar between M1 and M2.

3.1.4 DS4: Nominal scenario, new shape of B1, with B2

It can be seen from Table 10 and from Fig. 8 that the presence of B2 does not affect the results, as they remain quite similar to the ones obtained for $DS3$. The error distribution on the range estimation is biased towards positive values for M1 ($\mu\{\varepsilon_\rho\} = 1.19\ km$, $\sigma\{\varepsilon_\rho\} = 1.37\ km$), with a higher accuracy with respect to M2 ($\mu\{\varepsilon_\rho\} = 2.79\ km$, $\sigma\{\varepsilon_\rho\} = 1.51\ km$). For the estimation of the position of the COM of B1, M1 performs slightly better than M2, with a $\mu\{\varepsilon^n_{CoM}\} = 20.57\ px$ for M1 and $\mu\{\varepsilon^n_{CoM}\} = 20.70\ px$ for M2 and with a standard deviation of $\sigma\{\varepsilon^n_{CoM}\} = 16.08\ px$ and $\sigma\{\varepsilon^n_{CoM}\} = 18.14\ px$ respectively.

To conclude, both M1 and M2 are essentially unaffected by the presence of B2 in the images, as negligible differences are observed with respect to $DS3$ of the mean values of ε_ρ and ε^n_{CoM} are negligible.

3.2 $DS5$: Nominal scenario, different noises

Table 11 reports in a synthetic form the values of the mean and standard deviation of ε_ρ and ε^n_{CoM} for the images of $DS5$. Since the interest is to quantify the error introduced by the different noises, the absolute value of ε_ρ is analyzed, differently from the previous cases. The values of the mean are also represented in Fig. 9 for a simpler visualization.

It can be seen that in the overwhelming majority of the cases, both M1 and M2 performance suffer degradation when extra noise is added to the images. Both data-driven methods' performances worsen with the conditions of high blur, for both the estimations of the centroid of B1 and the range from it. The highest values of the means of $|\varepsilon_\rho|$ of M1 and M2 are $3.42\ km$ (BGL) and $1.96\ km$ (BgL) respectively, with a standard deviation of $2.43\ km$ and $1.85\ km$ respectively. The worst performance on the

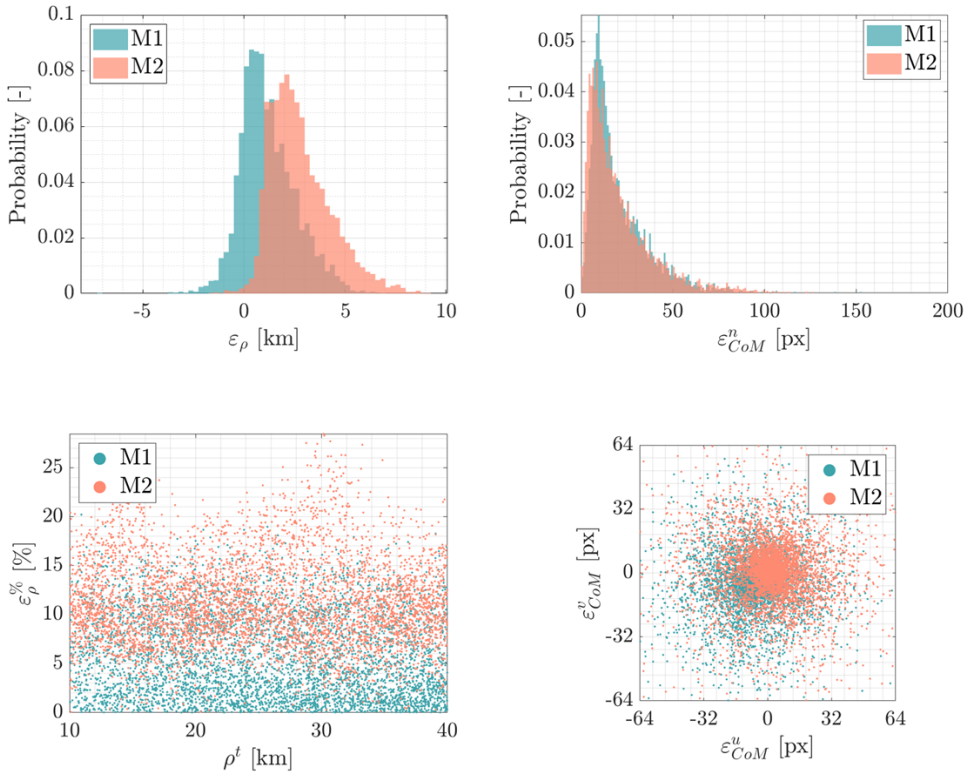

Figure 8: Range (left column) and centroiding (right column) error results of M1 and M2 for $DS4$.

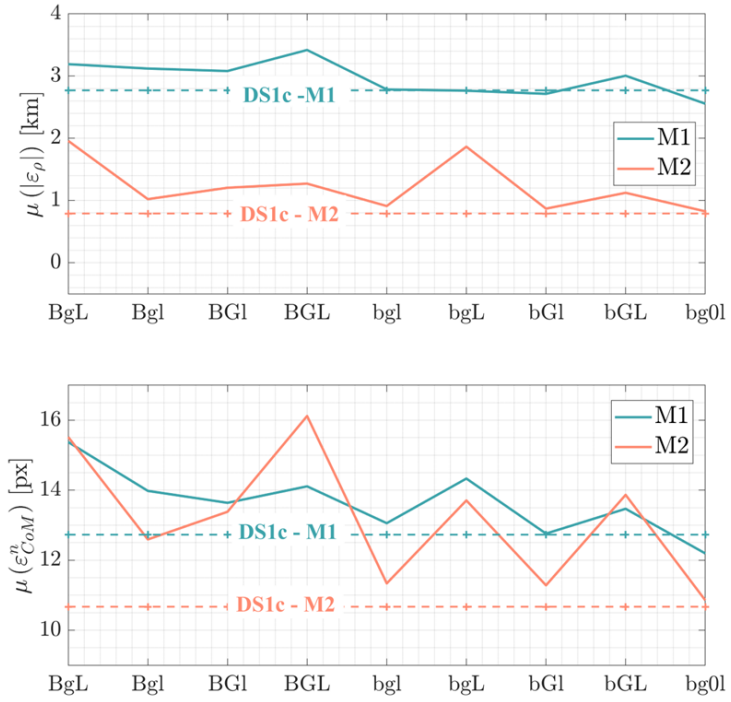

Figure 9: Range (top) and centroiding (bottom) error results of M1 and M2 for $DS5$.

Dataset	**IP**	$\mu(\sigma)\{\varepsilon^n_{CoM}\}[px]$	$\mu(\sigma)\{\lvert\varepsilon_\rho\rvert\}[km]$
BgL	M1	15.37 (11.38)	3.19 (2.17)
BgL	M2	15.52 (9.60)	1.96 (1.85)
Bgl	M1	13.98 (9.13)	3.12 (1.91)
Bgl	M2	12.59 (8.52)	1.02 (1.08)
BGl	M1	13.64 (8.67)	3.08 (2.09)
BGl	M2	13.38 (10.26)	1.20 (1.29)
BGL	M1	14.11 (9.29)	3.42 (2.43)
BGL	M2	16.12 (11.78)	1.27 (1.26)
bgl	M1	13.06 (8.69)	2.78 (1.64)
bgl	M2	11.34 (7.94)	0.91 (0.96)
bgL	M1	14.33 (10.25)	2.76 (1.84)
bgL	M2	13.71 (8.90)	1.86 (1.75)
bGl	M1	12.76 (8.03)	2.71 (1.84)
bGl	M2	11.28 (9.43)	0.87 (0.88)
bGL	M1	13.47 (8.54)	3.01 (2.06)
bGL	M2	13.87 (10.95)	1.12 (1.10)
bg0l	M1	12.19 (6.98)	2.56 (1.59)
bg0l	M2	10.86 (8.18)	0.83 (0.86)

Table 11: Performances of M1 and M2 on the estimation the position of the COM of B1 and the range from B1 for the different conditions of blur, gamma and mean and variance of the images of $DS5$.

centroiding estimation are given with conditions of BgL for M1, with a mean value of ε^n_{CoM} of $15.37\ px$ and a standard deviation of $11.38\ px$, and with conditions of BGL for M2, with a mean value of ε^n_{CoM} of $16.12\ px$ and a standard deviation of $11.78\ px$.

Fig. 9 shows also the values of $|\varepsilon_\rho|$ and ε^n_{CoM} obtained with the nominal scenario of $DS1_c$, as a reference to assess the robustness of the two algorithms with respect to the noises introduced with $DS5$. The performance of M1, when extra noise is added to the images, is consistent with that in $DS1_c$ while M2 seems more susceptible.

Finally, it is interesting to note that higher blur is negatively affecting both networks. Higher blur levels could challenge the convolutional kernels obtained during training, spreading the input image content across larger image patches. This result could inspire future data-driven IP designers to include a higher blur level in the training dataset in those cases in which robustness to noise is more demanding.

3.3 $DS7$ and $DS9$: DCP, new shape

Fig. 10 and Fig. 11 show the results on $|\varepsilon_\rho|$ and ε^n_{CoM} on the testing datasets $DS7$ and $DS9$ obtained by fine-tuning M1 and M2 using an incremental batch of images (8, 16,

32, 64, 128, 256, 512 and 1024) from the ECP phase represented in $DS6$ and $DS8$. The results obtained with M1 and M2 prior to the fine-tuning are also shown, represented by a batch size of 0 images. The pair $DS6 - DS7$ is expected to stress the difference in shape while the pair $DS8 - DS9$ stresses a greater shape difference, exacerbated in this setup also by the different rendering software used to generate the datasets (i.e. PANGU instead of CORTO). The aim of this analysis is to find the amount of images taken during the ECP that minimizes the estimation errors during the DCP, i.e. the number of images that tunes the two models' parameters to adapt their solution to a target different from the one seen during training. This scenario is especially relevant for small body missions, since the property of the target body are rarely precisely known before arrival.

Ideally, using a higher amount of images of the ECP for fine-tuning is expected to improve performances, as the data-driven models would have more information about the real target, minimizing possible discrepancies due to different modeling of the body's shape. However, as illustrated in this section, this scenario is more challenging than anticipated.

Considering the $DS6 - DS7$ case, observing the trends reported in Fig. 10, the number of images used for the fine-tuning of M1 and M2 is not improving performances on $DS7$ apart from the range estimate in M1.

Also, it is noted that all versions of fine-tuned M1 and M2 are performing worse than the untrained versions. Considering the $DS8 - DS9$ case, similar conclusions are obtained. Observing the trends reported in Fig. 11, the number of images used for the fine-tuning of M1 and M2 is not improving performances on $DS9$ apart from the centroid estimate in M2. The centroid estimate of M1 also seems to improve by increasing the number of images up to 256, after which it suffers a sudden drop in performance.

Lastly, it is also noted that both M1 and M2 have robust performances prior to fine-tuning, even in the case when different artificial environments are used (CORTO and PANGU) to generate training images. In particular, it can be seen from Fig. 11 that M1 performs better than M2, despite its worse performance with the other testing datasets generated using the same image generation tool of the training. However, it is also noted that M1 suffers in this dataset from the highest drop in convergence rates compared to all other scenarios presented in this work, as illustrated from the rates in Table 9.

A key conclusion is drawn from these results. Firstly, apart from specific cases (M1's range in DS7 and M2's centroid in DS9) in the majority of the cases considered, training episodes from the ECP dataset are drastically degrading both M1 and M2 performance for a robust application on the DCP datasets. A possible explanation of this phenomenon is given by the fact that both $DS1_a$ and $DS1_b$ used for the initial training are generated with a distance from the target that varies from $10\ km$ to $40\ km$. The images from ECP that are used for fine-tuning represent a new target shape at a distance between $20\ km$ and $30\ km$. Lastly, the testing dataset shows the new target shape at a range between $9\ km$ and $20\ km$. These different range intervals may explain why the $DS1$ dataset is more representative of the entire mission conditions than those encountered in ECP. Therefore, a possible explanation for the trend observed is that both data-driven methods are performing better when not fine-tuned, as their parameters were already optimized for ranges including the ones shown with $DS7$ and $DS9$. Effectively, both methods are challenged by the new training datasets of $DS6$ and $DS8$ to learn two contradicting information: learn about the new shape, and do that in a limited range interval. Even in an ideal training instance, in which a network would have specialized on the new shape but in a limited range interval, poor generalization performance would have been observed in a different geometric regime. This poses an interesting challenge in terms of fine-tuning data-driven methods for small-body applications. Fine-tuning episodes should be carefully designed to introduce the desired effect in the final networks.

4 Conclusion

This paper analyzes the robustness of two data-driven based IP algorithms designed to address the issue of centroiding (geometrical center) and range estimation of a binary asteroid system. The study examines the Hera mission's target proximity operations around the Didymos system as a case scenario. The data-driven methods are based on two different CNN approaches: M1, a heavier method based on the HRNet architecture and the poin regression strategy, and M2, a custom-built end-to-end architecture made of subsequent dilated convolutional layers.

The test campaign includes FTs with images generated with CORTO and PANGU, representing conditions different from the ones seen during training. In particular,

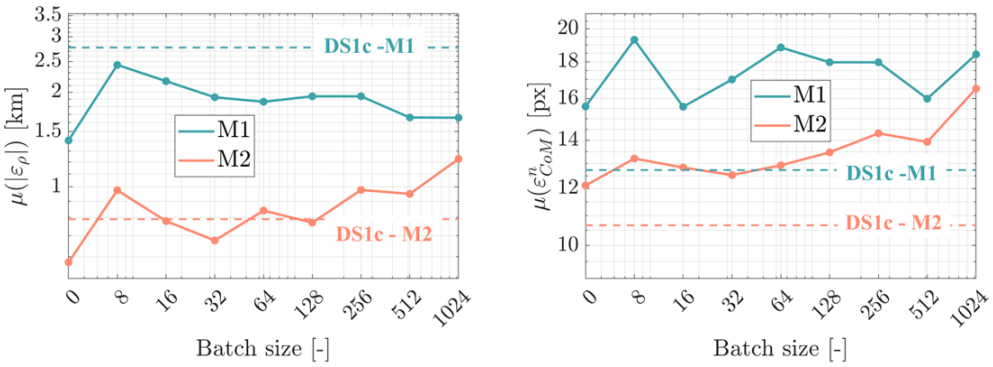

Figure 10: Range (left) and centroiding (right) error results of M1 and M2 for $DS7$.

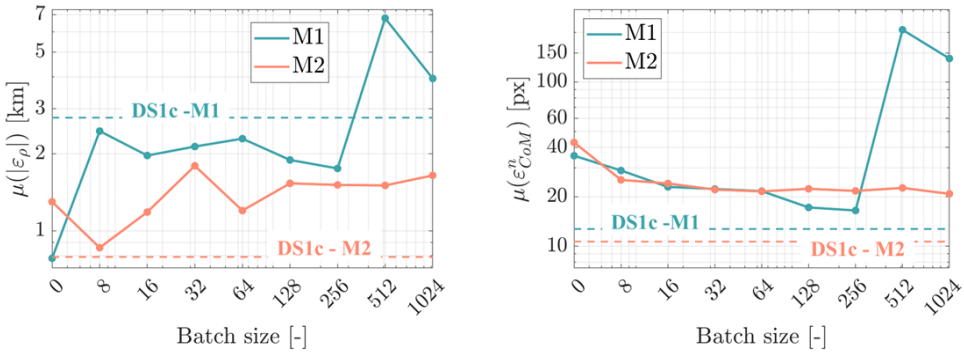

Figure 11: Range (left) and centroiding (right) error results of M1 and M2 for $DS9$.

the two data-driven methods are stress-tested with images showing the target asteroid system in conditions of high Sun phase angles, different noise levels, and different shape models of the target. Considering overall performances, the results show that neither method is affected by the presence of the secondary, that M2 is slightly more accurate and precise in the estimation of the centroid and range, but also shows greater sensibility to noises compared to the more robust M1.

Lastly, a potential operational scenario applicable to any data-driven method is also analyzed, considering a series of fine-tuning sessions simulating different batches of images downlinked to ground from the ECP phase of the Hera mission. This analysis turned out more complex and counterintuitive than expected, highlighting the challenges in fine-tuning these algorithms on specific characteristics of the target body. Indeed, both methods performed better in the DCP phase when fine-tuning was not applied, suggesting that both networks' training is affected more by different ranges than different shape models of the target asteroid. A possible explanation is that given the reference trajectories and the asteroid's rotation, both data-driven methods are trained considering images showing different orientations of the target with respect to the spacecraft. Therefore, the two methodologies are already trained with images showing a target with different shapes, since Didymos' irregular shape is shown differently in each image. On the other hand, the range depends on the apparent size of the asteroid in the images, which is unique at different range intervals.

To conclude, this work demonstrates that different data-driven approaches possess subtle differences that a coherent validation campaign can highlight. This ultimately represents an important tool for a mission designer to make informed decisions on their use. Considering overall performance, robustness, computational time, and fine-tuning inertia, one network was not clearly better than another, both exhibiting different sets of strengths and weaknesses that resonate for the better or for the worse in different untested conditions. M1 is less sensible to new untested conditions due to its larger capacity (with $\sim$8 times more parameters than M2), showing greater robustness. This comes at a cost of high inertia to fine-tuning, slower computational time, and some highly variable con-

vergence rates due to unclear heatmaps. M2 is a more efficient network, exhibiting greater accuracy and precision, shorter computational times, more agile fine-tuning, and always convergent to a solution. This comes at the cost of an increased sensibility to untested conditions.

Finally, the analysis illustrated in this work would have not been possible without a shared testbench serving as a common baseline. This also motivates the decision of the authors to make both datasets and results publicly available to encourage other researchers to propose different approaches. Future works could be focused specifically on the fine-tuning campaign of the ECP, including data-augmentation and image-manipulations as pre-processing steps of the ECP dataset. Furthermore, a hybrid training strategy and different architectures could be investigated.

Funding Sources

This study is co-funded and supported by the European Space Agency, The Netherlands under the Open Space Innovation Platform (ESA Contract No.4000133649/20/NL/MH/hm) and supported by GMV Defence and Space, Spain.

Acknowledgments

The authors would like to acknowledge the support of the Aerospace Centre of Excellence of University of Strathclyde, United Kingdom.

References

[1] G. Flandin, B. Polle, J. Lheritier, and P. Vidal, "Vision based navigation for autonomous space exploration," in *2010 NASA/ESA Conference on Adaptive Hardware and Systems, AHS 2010*. IEEE, 2010, pp. 9–16.

[2] T. Hashimoto, T. Kubota, J. Kawaguchi, M. Uo, K. Shirakawa, T. Kominato, and H. Morita, "Vision-based guidance, navigation, and control of Hayabusa spacecraft - Lessons learned from real operation," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 18, no. PART 1. IFAC, 2010, pp. 259–264.

[3] D. A. Lorenz, R. Olds, A. May, C. Mario, M. E. Perry, E. E. Palmer, and M. Daly, "Lessons learned from OSIRIS-REx autonomous navigation using natural feature tracking," *IEEE Aerospace Conference Proceedings*, pp. 1–12, 2017.

[4] A. Pellacani, P. Kicman, M. Suatoni, M. Casasco, J. Gil, and I. Carnelli, "Design and validation of a GNC system for missions to asteroids: the AIM scenario," *CEAS Space Journal*, vol. 10, no. 4, pp. 555–566, 2018.

[5] A. Pellacani, M. Graziano, and M. Suatoni, "Design , Development , Validation and Verification of GNC technologies," in *8th European Conference for Aeronautics and Space Sciences (EUCASS)*, 2019.

[6] P. Kicman, J. Lisowski, and A. Bidaux-Sokolowski, "Vision-based navigation around small bodies," in *Astrodynamics Network AstroNet-II*, G. Gómez and J. J. Masdemont, Eds. Cham: Springer International Publishing, 2016, pp. 137–149.

[7] G. Napolano, C. Vela, A. Nocerino, R. Opromolla, and M. Grassi, "A multi-sensor optical relative navigation system for small satellite servicing," *Acta Astronautica*, vol. 207, no. February, pp. 167–192, 2023.

[8] R. Brochard, J. Lebreton, C. Robin, K. Kanani, G. Jonniaux, A. Masson, N. Despré, and A. Berjaoui, "Scientific image rendering for space scenes with the surrender software," 2018.

[9] M. Pugliatti, C. Buonagura, and F. Topputo, "Corto: The celestial object rendering tool at dart lab," *Sensors*, vol. 23, p. 9595, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/23/23/9595

[10] M. Pajusalu, I. Iakubivskyi, G. J. Schwarzkopf, O. Knuuttila, T. Väisänen, M. Bührer, M. F. Palos, H. Teras, G. Le Bonhomme, J. Praks, and A. Slavinskis, "SISPO: Space Imaging Simulator for Proximity Operations," *PLoS ONE*, vol. 17, no. 3 March, pp. 1–22, 2022.

[11] H. Benninghoff, F. Rems, and T. Boge, "Development and hardware-in-the-loop test of a guidance , navigation and control system for on-orbit servicing," *Acta Astronautica*, vol. 102, pp. 67–80, 2014.

[12] F. Pace, E. Paolini, F. Sanfedino, D. Alazard, A. Colagrossi, V. Pesce, and S. Silvestrini, "GNC verification and validation," in *Modern Spacecraft Guidance, Navigation, and Control*. Elsevier Inc., 2023, pp. 647–684.

[13] M. Samaan, S. Lockhart, G. Holt, H. Mamich, S. Lockhart, G. Holt, and H. Mamich, "On-Ground Calibration and Optical Alignment for the Orion Optical Navigation Camera," in *John Junkins Dynamical Systems Symposium*, 2018.

[14] T. H. Park, J. Bosse, and S. D'Amico, "Robotic Testbed for Rendezvous and Optical Navigation: Multi-Source Calibration and Machine Learning Use Cases," in *AAS*, 2021, pp. 1–20. [Online]. Available: http://arxiv.org/abs/2108.05529

[15] A. Kaluthantrige, J. Feng, and J. Gil-Fernández, "Convolutional Neural Network-based autonomous navigation of Hera mission around Didymos," *Journal of Guidance, Control, and Dynamics*, pp. 1–29, (unpublished).

[16] M. Pugliatti, V. Franzese, and F. Topputo, "Data-driven image processing for onboard optical navigation around a binary asteroid," *Journal of Spacecraft and Rockets*, vol. 59, no. 3, pp. 943–959, 2022. [Online]. Available: https://doi.org/10.2514/1.A35213

[17] P. Michel, M. Küppers, A. C. Bagatin, B. Carry, S. Charnoz, J. de Leon, A. Fitzsimmons, P. Gordo, S. F. Green, A. Hérique, M. Juzi, Ö. Karatekin, T. Kohout, M. Lazzarin, N. Murdoch, T. Okada, E. Palomba, P. Pravec, C. Snodgrass, P. Tortora, K. Tsiganis, S. Ulamec, J. B. Vincent, K. Wünnemann, Y. Zhang, S. D. Raducan, E. Dotto, N. Chabot, A. F. Cheng, A. Rivkin, O. Barnouin, C. Ernst, A. Stickle, D. C. Richardson, C. Thomas, M. Arakawa, H. Miyamoto, A. Nakamura, S. Sugita, M. Yoshikawa, P. Abell, E. Asphaug, R. L. Ballouz, W. F. Bottke, D. S. Lauretta, K. J. Walsh, P. Martino, and I. Carnelli, "The ESA Hera Mission: Detailed Characterization of the DART Impact Outcome and of the Binary Asteroid (65803) Didymos," *Planetary Science Journal*, vol. 3, no. 7, p. 160, 2022.

[18] A. Pellacani, M. Graziano, M. Fittock, J. Gil, and I. Carnelli, "HERA vision based GNC and autonomy," *European Conference for AeroSpace Sciences*, pp. 1–14, 2019.

[19] N. G. Dias, P. Gordo, H. Onderwater, R. Melicio, and A. Amorim, "Analysis on the Isostatic Bipod Mounts for the HERA Mission LIDAR," *Applied Sciences*, vol. 12, no. 7, 2022.

[20] D. Izzo, M. Märtens, and B. Pan, "A survey on artificial intelligence trends in spacecraft guidance dynamics and control," *Astrodynamics*, vol. 3, no. 4, pp. 287–299, 2019.

[21] S. Sharma and S. D'amico, "Pose estimation for non-cooperative spacecraft rendezvous using neural networks," *Advances in the Astronautical Sciences*, vol. 168, pp. 3527–3546, 2019.

[22] A. Escalante Lopez, P. Ghiglino, and M. Sanjurjo-Rivo, "Applying Machine Learning Techniques for Optical Relative Navigation in Planetary Missions," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–11, 2024.

[23] P. Peñarroya, S. Centuori, M. Sanjurjo, and P. Hermosín, "A LiDAR-less approach to autonomous hazard detection and avoidance systems based on semantic segmentation," *Celestial Mechanics and Dynamical Astronomy*, vol. 135, no. 3, pp. 1–26, 2023. [Online]. Available: https://doi.org/10.1007/s10569-023-10140-9

[24] P. Arcaini, A. Bombarda, S. Bonfanti, and A. Gargantini, "Dealing with Robustness of Convolutional Neural Networks for Image Classification," in *2020 IEEE International Conference on Artificial Intelligence Testing (AITest)*, 2020, pp. 7–14.

[25] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–11.

[26] S. Dodge and L. Karam, "Understanding how image quality affects deep neural networks," in *2016 8th International Conference on Quality of Multimedia Experience, QoMEX 2016*, 2016.

[27] S. Ghosh, R. Shet, P. Amon, A. Hutter, and A. Kaup, "Robustness of Deep Convolutional Neural Networks for Image Degradations," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2018-April, 2018, pp. 2916–2920.

[28] F. Ferrari, V. Franzese, M. Pugliatti, C. Giordano, and F. Topputo, "Trajectory Options for Hera ' s Milani CubeSat Around (65803) Didymos," *The Journal of the Astronautical Sciences*, pp. 973–994, 2021.

[29] ——, "Preliminary mission profile of Hera's Milani CubeSat," *Advances in Space Research*, vol. 67,

no. 6, pp. 2010–2029, 2021. [Online]. Available: https://doi.org/10.1016/j.asr.2020.12.034

[30] P. Michel, A. F. Cheng, and M. Küppers, "Asteroid Impact and Deflection Assessment (AIDA) mission: science investigaation of a binary system and mitigation test," in *European Planetary Science Congress 2015, held 27 September-2 October, 2015 in Nantes, France, Online at http://meetingorganizer. copernicus. org/EPSC2015, id. EPSC2015-123*, vol. 10, 2015, p. 123. [Online]. Available: http://meetingorganizer.copernicus.org/EPSC2015/EPSC2015-123.pdf

[31] E. Smith, S. Zhan, E. Adams, M. Chen, D. Bekker, D. Carrelli, A. Johnson, C. Heistand, J. Thomas, A. Badger, L. M. Rodriguez, and M. Q. Tran, "Testing Early and Often: End-to-End Testing on the Double Asteroid Redirection Test (DART)," *IEEE Aerospace Conference Proceedings*, pp. 1–9, 2020.

[32] John Hopkins University Applied Physics Laboratory, "Design Reference Asteroid," Tech. Rep. 101955, 2022.

[33] ESA Headquarters, "HERA Didymos reference model," Tech. Rep. 5, 2021.

[34] P. Pravec, A. Meyer, P. Scheirich, D. Scheeres, C. Benson, and H. Agrusa, "Rotational lightcurves of Dimorphos and constraints on its post-DART impact spin state," *Icarus*, vol. 418, no. May, p. 116138, 2024. [Online]. Available: https://doi.org/10.1016/j.icarus.2024.116138

[35] ESA, "Hera Mission Instruments," 2021. [Online]. Available: https://www.heramission.space/hera-instruments

[36] H. Sierks, H. U. Keller, R. Jaumann, H. Michalik, T. Behnke, F. Bubenhagen, I. Büttner, U. Carsenty, U. Christensen, R. Enge, B. Fiethe, P. Gutiérrez Marqués, H. Hartwig, H. Krüger, W. Kühne, T. Maue, S. Mottola, A. Nathues, K. U. Reiche, M. L. Richards, T. Roatsch, S. E. Schröder, I. Szemerey, and M. Tschentscher, "The Dawn framing camera," *Space Science Reviews*, vol. 163, no. 1-4, pp. 263–327, 2011.

[37] F. Cabral and V. Moreno Villa, "Mission Analysis Report," GMV, Tech. Rep., 2022.

[38] ESA ESTEC, "HERA : Proximity Operation Guidelines," Tech. Rep., 2020.

[39] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2019-June, pp. 5686–5696, 2019.

[40] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-Resolution Representations for Labeling Pixels and Regions," 2019. [Online]. Available: http://arxiv.org/abs/1904.04514

[41] A. Kaluthantrige, J. Feng, and J. Gil-fernández, "CNN-based Image Processing algorithm for autonomous optical navigation of Hera mission to the binary asteroid Didymos," *Acta Astronautica*, vol. 211, no. May, pp. 60–75, 2023.

[42] M. Pugliatti, "Data-driven image processing for enhanced vision-based applications around small bodies with machine learning," Ph.D. dissertation, Politecnico di Milano, 2023.