POSE ESTIMATION OF DIDYMOS' MOON USING CNN-BASED IMAGE PROCESSING ALGORITHM FOR HERA MISSION Aurelio Kaluthantrige[1], J. Feng[1], J. Gil-Fernández[2], [1]University of Strathclyde, Glasgow, G1 1XJ, Scotland; mewantha.kaluthantrige-don@strath.ac.uk; [2]ESA/ESTEC, Noordwijk, 2200 AG, The Netherlands

Keywords: *Autonomous navigation, Pose estimation, Neural Network, Image Processing, Appearance-based model*

Introduction

The Asteroid Impact and Deflection Assessment (AIDA) is an international collaboration between the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA) aiming to investigate the binary asteroid system (65803) Didymos and to demonstrate asteroid deflection technique with kinetic impact. NASA's Double Asteroid Redirection Test (DART) mission successfully impacted Dimorphos, the moon of the binary system, in September 2022. ESA's contribution to AIDA is the Hera mission that will rendezvous with Didymos and observe the impact effects closely [1, 2]. Table 1 illustrates relevant properties of Didymos and Dimorphos measured directly through ground observations and updated with data collected with the on-board camera of the DART mission spacecraft [3, 4].

Table 1: Didymos' system properties [3, 4]

Parameter	Dydmos	Dimorphos
Gravitational parameter $[km^3/s^2]$	$3.5 \cdot 10^{-8}$	$2 \cdot 10^{-10}$
Extent along principal axes x $[m]$	849	177
Extent along principal axes y $[m]$	851	174
Extent along principal axes y $[m]$	620	116
Obliquity of the binary orbit with Ecliptic plane	$169.2°$	$169.2°$

The proximity operations of Hera consist of different phases that depend on the mission objectives. The Close Observation Phase (COP) is the proximity operation of Hera mission with the objective of obtaining high-resolution images of Dimorphos and fully characterizing the impact crater. Two CubeSats, Milani and Juventas, are planned to be released to perform detailed scientific observations and landing on Dimorphos. The success of the landing phase require accurate knowledge of the relative attitude of the secondary, as well as a high-level of autonomy to operate safely at close distances.

Autonomous optical navigation is designed for this phase based on line-of-sight and range measurements from both the primary body and Dimorphos in order to estimate the relative position of the spacecraft. This system includes the on-board Asteroid Framing Camera (AFC) taking images of the asteroid, an IP algorithm that extracts information from these images, and a navigation filter that processes the visual data and the dynamical environment. The close distance between the primary and the spacecraft during the COP allows the implementation of feature tracking relative navigation to solve the primary's relative attitude. Nevertheless, the relative attitude of Dimorphos remains unsolved as this methodology requires closer distances in order to be able to detect relevant features, due to Dimorphos' reduced size [5]. Furthermore, studies on the attitude dynamics of Dimorphos consequently to the DART impact suggest that unstable tumbling is possible, which could potentially add further challenges for the safety of the close proximity operations [6].

In this work we develop an appearance-based methodology to estimate the continuous pose (attitude) of Dimorphos during the COP using a Convolutional Neural Networks (CNN)-based Image Processing (IP) algorithm. The key of the developed algorithm is to leverage the lit limb of the secondary to estimate the quaternions representing the orientation of Dimorphos with respect to the spacecraft. By relying on the shape of the target, the developed algorithm is not constrained on the prior knowledge of the spinning state of the target and on the visibility of features on the surface of the target.

Methodology

In this section, the methodology of this research is described. Figure 1 shows the main steps of the undertaken pipeline. The objective of this work is to demonstrate the applicability of the developed algorithm during the COP. We train the developed appearance-based model with data collected during the previous phases of the mission.

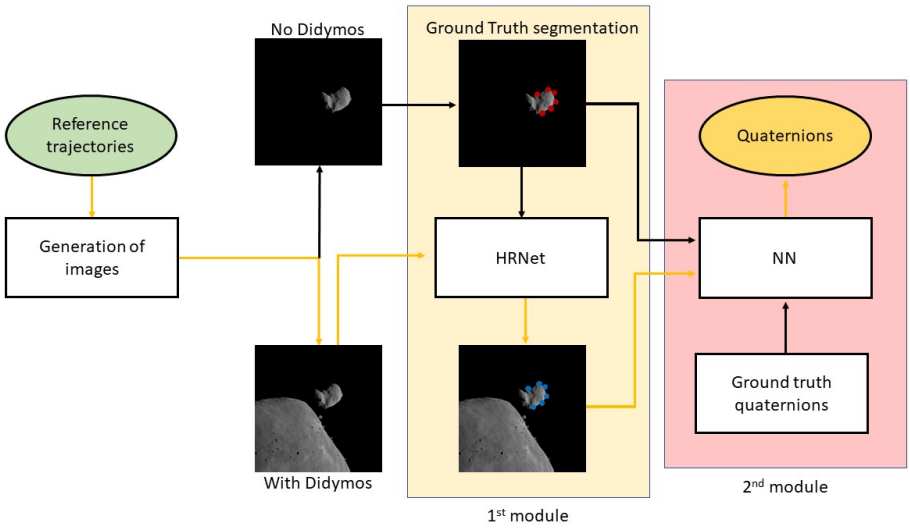

Figure 1: Pipeline of HRNet-based IP algorithm for Dimorphos' pose estimation.

The pipeline consists of two main modules: keypoint regression on the lit limb of Dimorphos, i.e. the segmentation, and its attitude estimation. The first module refers to the regression of a set of keypoints on the images of the target, and it is carried out by CNNs using the High Resolution Network (HRNet) architecture [7]. The second module refers to mapping the regressed keypoints to the quaternions describing the relative attitude of Dimorphos with respect to the camera, in order to have a correlation between the lit limb of Dimorphos and its pose. The second module is carried out by a Neural Network (NN) developed in this work.

We rely on synthetic images generated with the software Planet and Asteroid Natural scene Generation Utility (PANGU). PANGU is used to generate two sets of images, one with both bodies and one without the presence of Didymos. The latter set is processed to retrieve the Ground Truth (GT) pixel position of keypoints which are used to supervise the training and validation of the HRNet for the segmentation module. The keypoints are 31 points selected from the projected surface of Dimorphos on the image plane which include: the Center of Mass (COM) of Dimorphos and 30 points on its lit limb. The GT keypoints are used together with the GT quaternions of the attitude of Dimorphos to supervise the training of the NN.

The yellow arrows of Figure 1 illustrates the path the images captured by the AFC camera during the selected trajectory are processed by the algorithm during the test case scenario in order to estimate the desired output. Details of the main steps are described in the rest of this section.

Reference trajectories: The adopted reference frame is the Target Body Equatorial Inertial (TB), which has the origin located on Didymos, the X-axis pointing towards the vernal equinox, and the XY plane coplanar to the equatorial plane of Didymos. The relative motion of the Sun around Didymos is retrograde as the binary system's orbit obliquity with respect to the ecliptic plane is larger than $90°$, as shown in Table 1.

The COP trajectory is provided by ESA. Figure 2 illustrates the trajectory of the spacecraft, together with the position of the Sun (scaled down in the illustration) and the orbit of the secondary, all shown in the TB reference frame. The position of the Sun is calculated using the Jet Propulsion Laboratory Small Body Database [8]. The trajectory consists on several hyperbolic arcs with a total duration of 14 days. The only forces considered for each arc are the point mass gravitational attractions of both the primary and the secondary asteroids. Orbital manoeuvres are performed at the joint of two arcs.

The range from the primary varies between a minimum of $4\ km$ to $22\ km$. It can be seen from the XY view of Figure 2 that the COP trajectory is located in between the Sun and Didymos, in order to provide the AFC camera with bright images of both bodies for Line of Sight navigation.

The training and validation database of images is generated considering a previous phase of the mission, specifically the Early Characterization

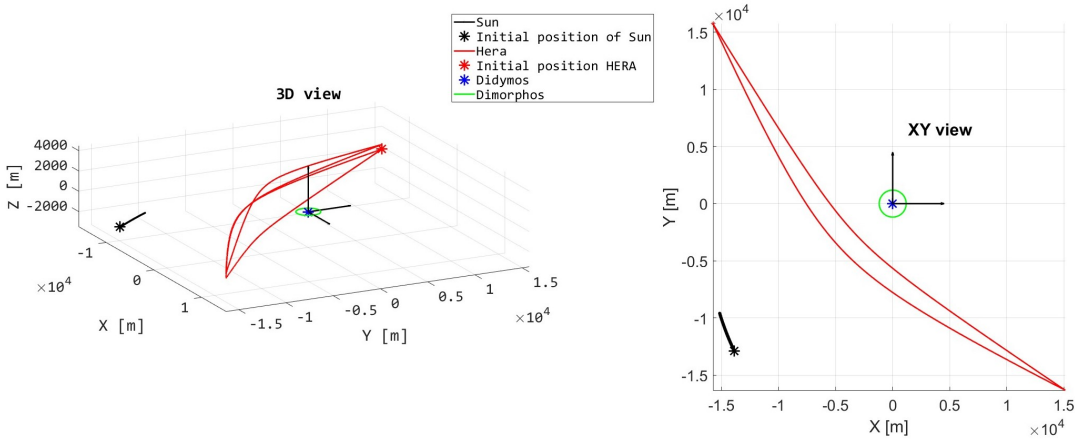

Figure 2: COP trajectory

Phase (ECP), whose trajectory is provided by ESA and it is shown in Figure 3. The trajectory consists of 4 hyperbolic arcs, with an initial epoch of $t_{in} = 9012\ days$ and a final epoch of $t_{fin} = 9026\ days$, calculated in the Modified Julian Date 2000. The only forces considered for each arc are the point mass gravitational attractions of both the primary and the secondary asteroids. Orbital manoeuvres are performed at the joint of two arcs. The durations of the 1^{st} and 3^{rd} arcs are both 4 days while the durations of the 2^{nd} and 4^{th} arcs are both 3 days. The range from the primary varies between a minimum of $20\ km$ and a maximum of $30\ km$.

Generation of images: The software PANGU is used to generate the database of images for this research. PANGU is a simulation tool that models planet and asteroids surfaces and provides a high-fidelity visualization of images while operating at near real-time speeds. The software has been developed by the STARDundee engineering company [9]. The models of Didymos, Dimorphos and the camera are provided by GMV and their shapes are updated with the data collected with the DART mission shown in Table 1. The software generates greyscale images detected by the camera and shows them on the PANGU viewer, which is a plane with the size of the image (shown in Table 2) and the origin of the coordinated frame set at the top left corner. The horizontal and the vertical axes of the plane are referred as i-direction and j-direction respectively. The flight file system of PANGU is operated in order to visualize the binary asteroid system during the trajectories. Flight files are the input to PANGU and they control the viewer to generate images taken at selected epochs of the reference trajectory in the TB reference system, considering the position of the Sun (range, Azimuth and Elevation) and the positions and the orientations (quaternions) of both the binary asteroid system and the AFC camera (joined with the spacecraft) [9].

Table 2: AFC properties [10, 11]

FOV	Focal Length: f	Aperture	Image size	Pixel Size: ν
$5.5°$	$10.6\ cm$	$2.5\ cm$	$1024 \times 1024\ pixels$	$14\ \mu m$

For asteroid imaging, the AFC has its boresight pointing towards the binary asteroid system and the vertical axis of the camera is perpendicular to the direction of the Sun with respect to the spacecraft [12]. PANGU adopts the boresight, the vertical and the horizontal axes of the camera respectively as the Z- the Y- and the X-axis of the camera reference frame. Therefore, the position vector of the Sun with respect to the spacecraft lies on the XZ plane of the camera frame. As a result, the images shown in the PANGU viewer always represent the binary system illuminated from the right side.

In this work, PANGU is used to generate:

- Dataset 1: 40000 images taken during the ECP trajectory and used for the training and validation of the HRNet and the NN; two fictitious additional arcs are considered, the first one connecting the end of the 2^{nd} arc with the beginning of the 1^{st} arc, the second one

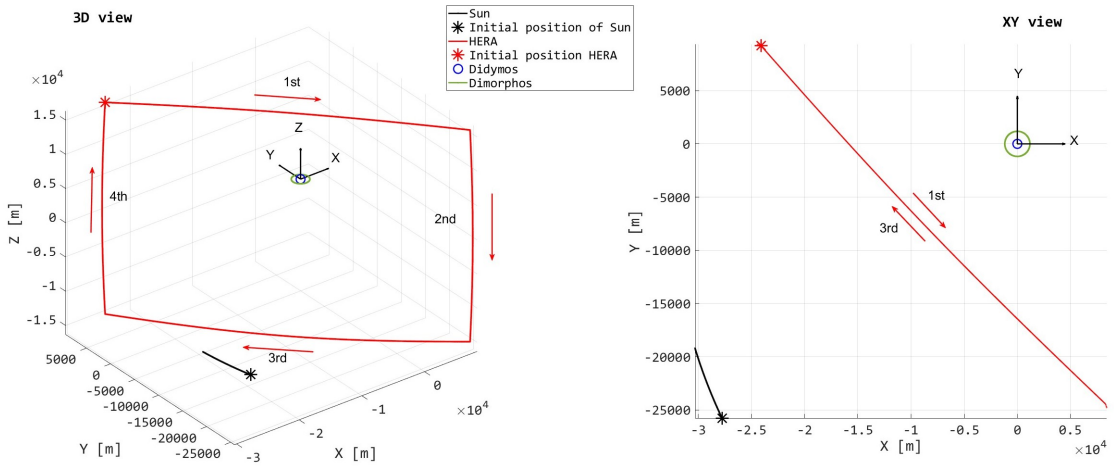

Figure 3: ECP trajectory

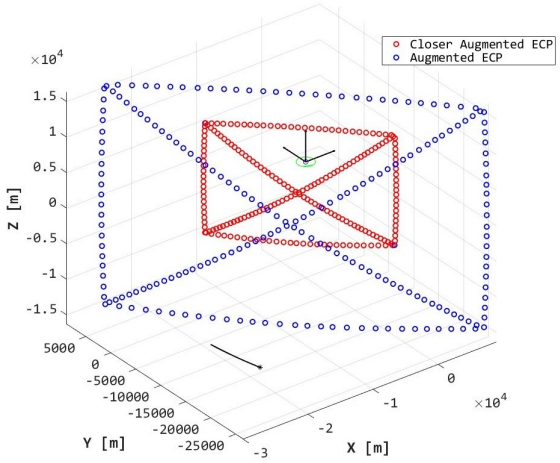

Figure 4: Augmented ECP trajectory

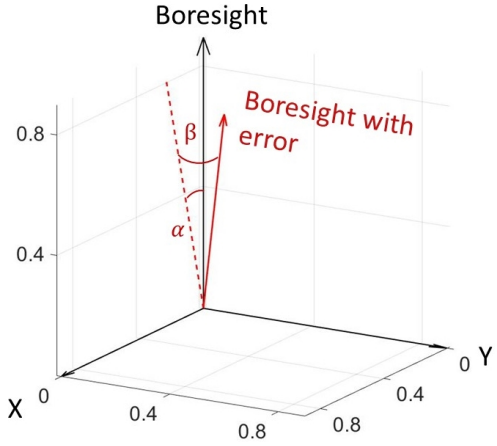

Figure 5: Camera pointing with error

connecting the end of the 3^{rd} arc with the beginning of the 1^{st} one, as shown in Figure 4. The augmented ECP trajectory is sampled randomly to generate a secondary trajectory closer to the target, with a minimum distance of $7\ km$, in order to provide the HRNet and NN with a training dataset of images showing the asteroid in multiple configurations;

- Dataset 2: 6052 images taken sampling the COP trajectory every $200\ seconds$ and used for the testing of the HRNet and the testing of the NN;

Each set of images is generated twice, one presenting both bodies and the other one without Didymos.

Ground Truth data: The GT data of the developed pose estimation algorithm consist in the COM, 30 segmentation keypoints on the lit limb and the relative rotation matrix of Dimorphos with respect to the camera expressed in quaternions.

When the camera is pointing perfectly towards the secondary, the latter are displayed in the middle of the PANGU viewer. With the conditions that the camera is pointing directly to the secondary, the Geometrical Centre (GC) of the secondary that is the arithmetic mean position of all the points belonging to the body, is located at the central pixel with the coordinates $(i, j) = (512, 512)\ pixels$ in the PANGU viewer. The COM of the secondary almost coincides with its GC because of its ellipsoidal shape. Since the images used in this work are all generated with PANGU, it is assumed that the GC of Dimorphos is its centroid. Training the

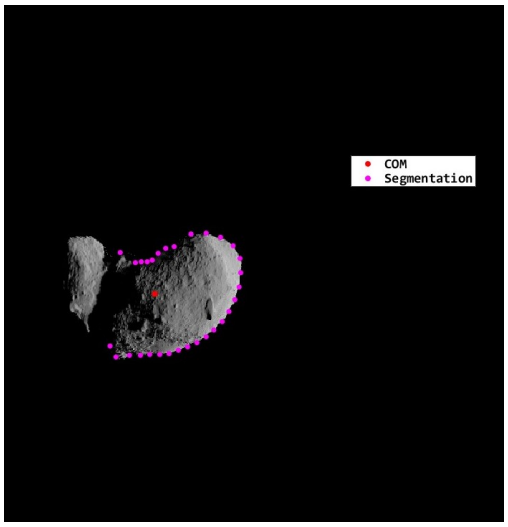

Figure 6: Example of GT keypoints on generated image with PANGU

CNN algorithm with a set of images with perfect pointing conditions will result in an issue of lacking label variability. To overcome this issue, a pointing error represented by spherical coordinates and defined by two angles α and β is introduced at each epoch of the trajectory in the boresight direction of the camera reference system, as shown in Fig. 5. As a result, the generated images are shifted from the central position of the PANGU viewer. In order to make sure that both bodies lie within the FOV of the AFC camera, random values within an interval of $[-1, 1]^\circ$ are considered for both α and β. With these values, the primary and secondary locations are shifted around in the PANGU viewer. By calculating the shift in pixels of the secondary from its central position, the GT pixel coordinates of the COM of Dimorphos is calculated for each value of α and β.

The 30 segmentation keypoints are selected within an angular aperture of 2θ with $\theta = 130^\circ$ considering that the asteroid is illuminated from the right side of the image plane. To generate the ground truth pixel position of the segmentation keypoints, Didymos is hidden from the images as its presence in front of Dimorphos or near its border would represent a disturbance. Figure. 6 shows an example of the GT COM and 30 keypoints on the lit limb of Dimorphos.

When the range of the spacecraft from Dimorphos decreases, the projected image of the target on the camera is bigger, and the keypoints of the lit limb are located further away from each other. In order to lose the dependency of the keypoints' positions with the range and, thus, enable the NN of the second module to solve for the relative attitude of Dimorphos independently from the size of Dimorphos in the image, the coordinates (X, Y) of each i-th keypoint on the lit limb is normalized using Eq. 1.

$$x_i = \frac{30(X_i - X_{COM})}{\sum_1^{30} |X_i - X_{COM}|}$$

$$y_i = \frac{30(Y_i - Y_{COM})}{\sum_1^{30} |Y_i - Y_{COM}|} \tag{1}$$

Given the rotation matrix R_T^C from the TB reference frame to the camera reference frame and the rotation matrix R_T^B from the TB reference frame to the body fixed frame of Dimorphos, the relative rotation matrix R_C^B that represents the attitude of Dimorphos with respect to the camera can be calculated with Eq. 2.

$$R_C^B = R_T^B \cdot inv(R_T^C) \tag{2}$$

The relative rotation matrix R_C^B is translated using Eq. 3 into the 4 quaternions, used to supervise the training of the NN.

$$\begin{aligned} q_0 &= \frac{1}{2}\sqrt{1 + C_{11} + C_{22} + C_{33}} \\ q_1 &= \frac{1}{4q_0}(C_{23} - C_{32}) \\ q_2 &= \frac{1}{4q_0}(C_{31} - C_{13}) \\ q_3 &= \frac{1}{4q_0}(C_{12} - C_{21}) \end{aligned} \tag{3}$$

The notation C_{ij} refers to the i-th and j-th element of the rotation matrix R_C^B. In this work, the short rotation around the Euler principal axis ($q_0 > 0$) is considered.

HRNet and NN: The HRNet architecture is shown in Fig. 7. The network maintains the high resolution representations of the input images by connecting multiple subnetworks in parallel. The first stage is a high-resolution subnetwork. New stages are formed from the gradual introduction of high-to-low subnetworks. To maintain the high-resolution representation, repeated multiscale fusions are performed using low-resolution representation of the same depth and level. The last high-resolution representation is then used for the regression of the selected visual data [13].

The keypoints to regress for each image of Dataset 1 and Dataset 2 are 31, which are the

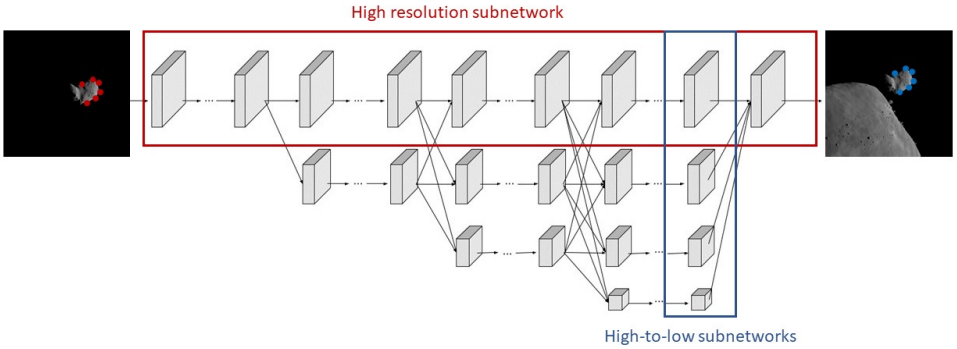

Figure 7: HRNet architecture

COM_{Dim} and the 30 points on the lit limb of Dimorphos. Each input image of the HRNet is coupled with the corresponding keypoints that are used to supervise the training to regress the keypoints locations on the testing dataset. For this work, the CNN architecture of the pose-hrnet-w32 that was previously implemented in [14] is used, where 32 represent the widths of the high-resolution subnetwork in the last three stages. During the training, the validation dataset is used beside the training one to compute the validation losses and avoid overfitting. The Adam optimizer is used with a cosine decaying learning rate with initial value of 10^{-3} and decaying factor of 0.1. The total parameters involved in the training process are $28,536,575$. The network is trained for 100 epochs on the virtual machine provided by Google Colab with the NVIDIA V100 Tensor Core GPU, and it takes around $48\ hours$.

The NN architecture consist on 7 fully connected layers with Rectified Linear Unit functions and a number of nodes per layer that decreases throughout the network. The input layer has as number of neurons equal to 60, which represents the x and y position of each of the 30 keypoints and normalized as mentioned in the previous section. The output layer has 4 neurons which are normalized in order to get the estimation of the 4 quaternions. The Adam optimizer is used with a cosine decaying learning rate with initial value of 10^{-3} and decaying factor of 0.1. The total parameters involved in the training process are $13,719,124$. The network is trained for 1000 epochs on the NVIDIA GeForce RTX 2070 with Max-Q Design, and it takes around $6\ hours$.

Didymos is not hidden in the images as done for the GT data, so that the HRNet and the NN are trained to regress the location of the keypoints and to subsequently solve for the pose of Dimorphos despite the disturbance introduced by the presence of Didymos. Nevertheless, some images of Dataset 1 and Dataset 2 are discarded because of the projection of Dimorphos on the camera being partially outside of the FOV or partially covered by Didymos. The input database of the HRNet and the NN consists of 33530 (83.13%) images for training (Dataset 1), 1790 (4.43%) images for validation (Dataset 1) and 5016 (12.44%) images for testing (Dataset 2).

Results

In this section, the results of the developed algorithm for the relative attitude estimation of Dimorphos with respect to the camera are presented. Firstly, the accuracy of the HRNet on the estimation of the positions of the keypoints on the images of the testing dataset is evaluated with the Root Mean Squared Error (RMSE) defined with Eq. 4:

$$RMSE_m = \sqrt{\frac{\sum_{n=1}^{N}(P_{mn}^{GT} - P_{mn}^{pred})^2}{N}} \tag{4}$$

where P_{mn} represents the m-th keypoint; the index n refers to the n-th image of the $N = 5016$ images of the testing dataset; $RMSE_m$ is the

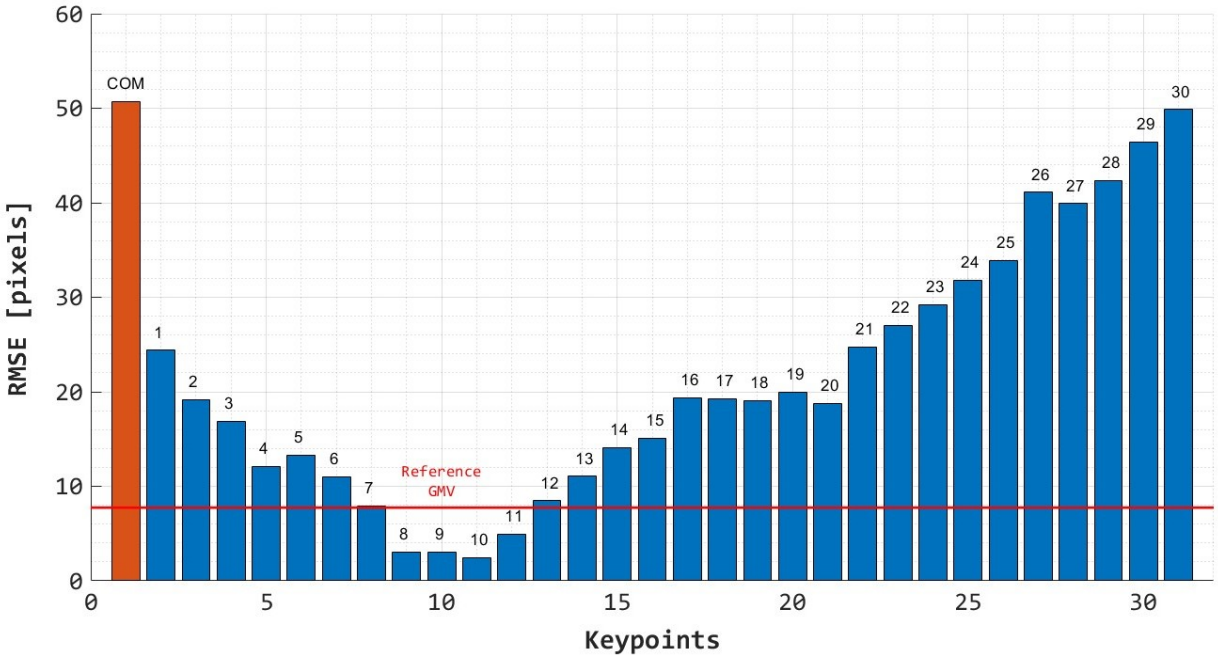

Figure 8: Accuracy of the keypoints regression step

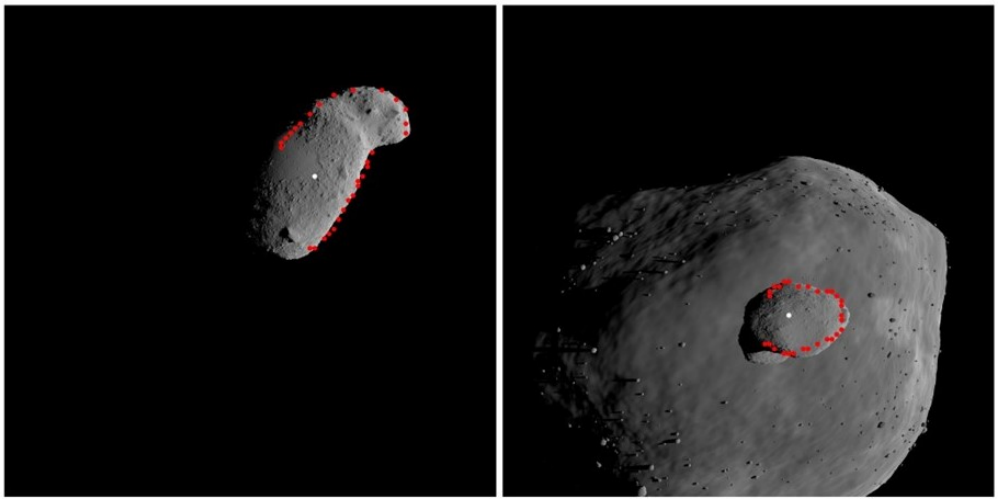

Figure 9: Keypoints regression results for two images of the testing dataset

RMSE between the GT position of the m-th keypoint P_{mn}^{GT} and the estimation P_{mn}^{pred} using the HR-Net. The $RMSE$ value of $7.746\ pixels$ obtained by the Maximum Correlation with a Lambertian Sphere (MCLS) IP algorithm developed by GMV to estimate the position of the centroid of Didymos during the ECP of Hera mission is given as a reference, since their estimation meets the pointing accuracy required by the mission [15]. This result is obtained by applying the algorithm over a set of 243 images of the 1st arc of the ECP generated with PANGU by GMV.

Further, the accuracy of the NN on the estimation of the quaternions representing the relative attitude of Dimorphos with respect to the camera is assessed with the percent error, defined in Eq. 5.

$$err_{q_i} = \frac{|q_i^{pred} - q_i^{GT}|}{q_i^{GT}} \cdot 100\% \tag{5}$$

Keypoints regression results: The accuracy of the HRnet on the estimation of the positions of the 31 keypoints is evaluated with the Root Mean Squared Error (RMSE) calculated for each point throughout the whole Dataset 2. Figure 8 is a bar chart of the obtained $RMSE$ values. The orange bar represents the $RMSE$ obtained for the COM of Dimorphos while the other bars represent the $RMSE$ obtained for the keypoints on the lit limb. It is observed that the worst performance of the HRnet is obtained for the COM, with a $RMSE$ value of $50.73\ pixels$. This is due to the fact that Dataset 2 contains images where Dimorphos is in partial or total eclipse because of the shadow of Didymos. In case of total eclipse the only visible part of Dimorphos is its limb, which results in a higher accuracy and a lower $RMSE$ obtained for the regression of the other points. The best performance of the keypoint regression is obtained for point 10, with a $RMSE$ value of $2.47\ pixels$.

The difference of the estimation accuracy among the keypoints depends on the different lighting conditions and on the shape of the asteroid. Nevertheless, the values of the $RMSE$ differ from the one obtained using the MCLS IP algorithm by a maximum of around $40\ pixels$, which is negligible compared to the size of the image. Figure 9 illustrates two sample images generated by PANGU during the COP, together with the keypoints estimated by the HRNet.

Estimated quaternions results: The results on the estimation of the quaternion representing Dimorphos relative attitude with respect to the camera over the 5016 images of the testing dataset is shown in Figure 10. It can be seen that in general the estimated quaternions follow the behaviour of their respective ground truth value. The percent error calculated with Eq. 5 is summarized in Table 3. The images where the percent error is lower than 20% are considered as cases where the attitude estimation algorithm has succeeded and they are shown in the last column of Table 3 with the success rate, defined as the percentage of successful estimations among the total. It is observed that the estimation of the q_0 is the worst with a mean value of the percent error of 76.12% and a success rate of 56.22%. This is due to the fact that in this work only the short rotations are considered ($q_0 > 0$), which is a constraint on the value of q_0 not considered in the implementation of the architecture of the NN. The minimum mean value of the percent error is obtained for q_3 with 12.68%, while the maximum success rate for q_1 with 85.31%. The successful cases are different for each quaternion.

Table 3: Percent error on estimated quaternions of testing dataset

	Mean [%]	**Minimum** [%]	σ [%]	**Success rate** [%]
q_0	76.12	2.13	15.02	56.22
q_1	16.59	0.01	24.69	85.31
q_2	13.81	0.005	21.52	70.18
q_3	12.68	0.001	19.75	69.06

The interception of the successful cases is 51.14% and it identifies the success rate of the developed algorithm for the whole relative attitude estimation of Dimorphos with respect to the camera.

The standard deviations of the percent errors are low for each quaternion, which means that 68.27% of the estimations have a percent error close to its mean value. Ultimately, the developed algorithm for the quaternion estimation of the relative attitude of Dimorphos with respect to the camera leveraging the position of keypoints regressed on the images is not ideal for a navigation filter, as the error is far from the Gaussian distribution [16].

Conclusion: This paper develops a CNN-based IP algorithm addressing the problem of attitude estimation of Dimorphos during the COP proximity operation of Hera around the target. The challenges tackled by the developed methodology include the potential unstable spinning state of the target and the limited visibility of features on the surface of the target due to its reduced size and the range from the spacecraft. The proposed pipeline is dived into two modules: the regression of 30 keypoints on the lit limb of Dimorphos together with the regression of its COM, run by the HRNet, and the attitude estimation, run by a custom-made NN.

The results show that the HRNet is able to estimate the position of the 30 keypoints and the COM with high accuracy. In particular the regression is robust to the presence of Didymos. The algorithm can provide an accurate estimation of the quaternions for 51.14% of the COP trajectory. To increase the success rate of the developed algorithm, the training database can be augmented using additional images generated with the Detailed Characterization Phase (DCP) trajectory or other fictitious trajectory segments.

Nevertheless, the proposed pipeline is applied to the current known models of Didymos and Dimorphos. Assuming that the same methodology is applied to images representing the actual shapes of

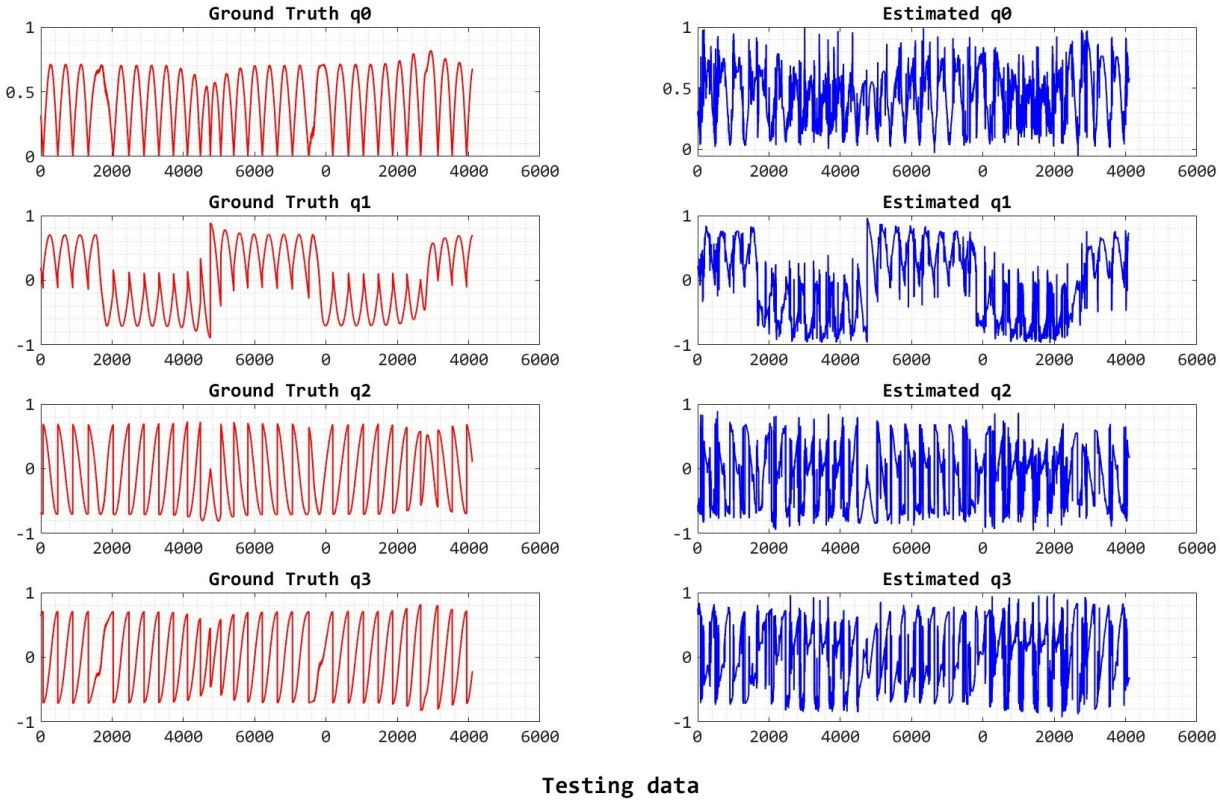

Figure 10: Ground Truth vs Estimated quaternions

Didymos and Dimorphos, the obtained results will be different. The regression of the 31 keypoints will be affected limitedly, as the shape of the target will be learned by the HRNet. On the other hand, the estimation of the quaternions depends on the shape of the target. Therefore, the NN will require a fine-tuning of its parameters with a dataset of images taken during the first days of the ECP and/or DCP in order to be able to map the keypoints with the quaternions.

Our developed methodology contributes to the Space Situational Awareness by improving the robustness and the autonomy of the navigation strategy of the first mission ever testing asteroid deflection. Specifically, the unique contribution represented by the attitude estimation of the target can be applied to any object in space, for instance space debris.

Future work would go into the direction of other applications for the CNN-based IP algorithm. For instance, an additional output useful for the navigation during the proximity operations is the complete pose (attitude and position). Subsequently, the algorithm will be incorporated with a navigation filter to estimate the state of the spacecraft and to quantitatively evaluate the improvement of navigation performance.

References:

[1] P. Michel, et al. (2018) The HERA mission, European component of the asteroid impact and deflection assessment (AIDA) mission to a binary asteroid.
[2] P. Michel, et al. (2015) Asteroid Impace and Deflection Assessment (AIDA) mission: science investigation of a binary system and mitigation test.
[3] A. P. Laboratory (2022) (101955):8. [4] ESA Headquarters (2021) HERA Didymos reference model *Tech. Rep. 5*.
[5] M. Küppers, et al. (2022) Hera mission requirements document *Tech. rep.*.
[6] H. F. Agrusa, et al. (2021) *Icarus* 370(July 2021):114624 ISSN 10902643 doi. arXiv:2107.07996.
[7] K. Sun, et al. (2019) *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June:5686 ISSN 10636919 doi.arXiv:1902.09212.
[8] NASA (2021) JPL Solar System Dynamics.
[9] D. university (2019) *Planet and Asteroid Natural*

Scene Generation Utility User Manual 4.
[10] ESA (2021) Hera mission instruments.
[11] H. Sierks, et al. (2011) *The Dawn framing camera* vol. 163 ISBN 1121401197454 doi.
[12] ESA Estec (2020) HERA: Proximity Operations Guidelines *Tech. Rep. 6*.
[13] K. Sun, et al. in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 5686–5696 june, 2019.
[14] K. Sun, et al. (2019) in *CVPR*.
[15] A. Pellacani, et al. (2019) in *European Conference for AeroSpace Sciences* 1–14.
[16] M. I. Ribeiro (2004) *Institute for Systems and Robotics Lisboa Portugal* (February):42.