AAS 21-383

NAVIGATION ABOUT IRREGULAR BODIES THROUGH SEGMENTATION MAPS

Mattia Pugliatti[*], Francesco Topputo[†]

Optical navigation about small-bodies can be performed at different scales and with different techniques during proximity operations. Traditional methods however are influenced by pixel intensity due to illumination conditions and often provide a navigation solution only when coupled with filtering techniques. In this work, a navigation method for small-body applications is presented that makes use of segmentation maps. By converting a grayscale image into its segmented equivalent the pixel content is highly reduced but at the same time its meaning is enriched since the pixel value is providing direct information on feature type and distribution across space. This is exploited in an autonomous navigation method in two steps. A Convolutional Neural Network is designed to generate a rough estimate of the position of a spacecraft in a small-body fixed reference frame, whose surrounding has been divided into 1176 classes. A Normalized Cross-Correlation technique is then applied to the reduced search space to generate a precise position estimate. The methodology proposed is trained and validated on a database of segmented synthetic maps of 49716 samples of Didymos and Hartley each, while a series of 5 scenarios are tested. The CNN is capable to predict the correct class with an accuracy of 75.94% and 68.60% respectively for Didymos and Hartley, while the overwhelming majority of the other cases are predicted just next to the correct classes. The CNN is robust to various illumination conditions, is capable to work outside the range of distances considered during training, performs well when predicted masks are used, and also selects independently the type of features to rely on for classification depending on the body. When coupled with NCC, a position estimate with a relative error below $5 - 8\%$ the range from the asteroid can be achieved.

INTRODUCTION

Missions toward small-bodies such as asteroids and comets are becoming increasingly interesting, pushed by scientific, planetary defense, and resource exploitation motivations [1]. These bodies are thought to enclose valuable information on the primordial state of the Solar System, which can be used to improve our knowledge of planetary systems formation. Their extensive presence in the Solar System poses both a threat to our planet and at the same time offers plenty of opportunities for resource exploitation or for technology demonstration.

This trend of increasing interest towards small-bodies has been mutually supported by a trend in miniaturization for deep-space applications [1, 2, 3], enabling low-cost and high-risk CubeSat missions. Standalone missions such as M-ARGO [4] and NEAScout [5] have been proposed to rendezvous with an asteroid. Various secondary payloads are being designed to visit Didymos

[*]PhD student, Department of Aerospace Science and Technology, Politecnico di Milano, mattia.pugliatti@polimi.it.
[†]Associate professor, Department of Aerospace Science and Technology, Politecnico di Milano, francesco.topputo@polimi.it.

system either for a shorter period of time such as for the LiciaCube mission [6] or for a longer period as for the case of the two Hera's CubeSats, Milani [7] and Juventas [8].

To cut operational costs and to increase spacecraft safety, missions toward small-bodies often incorporate autonomous capabilities. Because images can easily be available on-board the spacecraft with simple and low-cost solutions, these are often fundamental in enabling autonomy, especially regarding on-board state estimation. To tackle these new challenges an increasing variety of techniques have been developed in the framework of autonomous optical navigation about irregular bodies for proximity operations.

Center of Brightness (CoB) methods are simple, robust, and easy to use [9], but they lack accuracy when highly irregular shapes are considered or when high phase angles with the Sun are used due to large Center of Mass (CoM) offset. To overcome these limitations, modified versions that introduce corrective terms as a function of the illumination conditions that move the CoB towards the CoM can be adopted [10, 11].

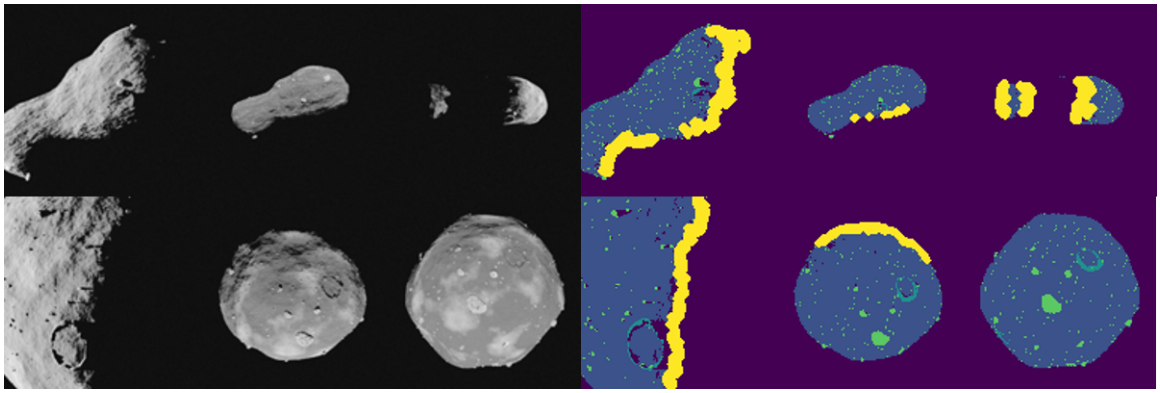

Figure 1: Examples of synthetic images and their segmentation maps of asteroid Didymos and comet Hartley considered in this work taken with 10° FOV at various ranges.

Centroid Apparent Diameter [12, 13] (CAD) techniques based on ellipse or limb fitting [11, 14] and Lambertian Sphere Correlation (LSC) [15] combine line-of-sight information with the body's real dimension knowledge and apparent size to generate a directional range. In CAD techniques a geometrical figure, such as a circle or an ellipse, is fit with the detected limb of the small-body. The center of the geometrical figure adopted is then used as an approximation of the object CoM while from simple scaling laws a comparison between the real and apparent diameter of the body is used to generate a range. A well-known non-iterative CAD method that works with tri-axial ellipsoids is described in [13]. LSC uses a spherical model as a proxy of the real object and tries to achieve a maximum correlation between the binarization of the real image and a library of Lambertian sphere templates [15]. These are either stored on-board or rendered online.

However, these techniques assume regularly shaped objects. Thus, they are often preferred for regular targets as planets and moons. Nonetheless, attempts have been made to extend them also for irregular objects, most notably [14]. CoB, CAD, and LSC are not able to generate alone a relative state with respect to an asteroid fixed reference frame, instead, they only provide line-of-sight information, eventually complemented with relative range. In order to extract the spacecraft position with respect to an asteroid fixed reference frame, a priori information about the state shall be provided and then filtered with optical and attitude observations.

Correlation methods can be used to overcome this limitation if the irregular shape model is known. The outline can be extracted as a feature and then matched with a pre-existing catalog [16] or the whole body shape can be binarized and cross-correlated with a library of shapes generated from different point of views [17]. Both methods could potentially provide the spacecraft state in an asteroid fixed reference frame, but they could suffer fatal ambiguities from opposite points of view as the outline may correlate with multiple peaks over opposite regions of the search space. To overcome this limitation, features extracted from grayscale images could complement the information provided by the outline alone.

All the aforementioned methods can only be applied if the body's outline is resolved in the image and fits within the FOV of the camera. At closer distances, features based methods are usually preferred. These methods can be relative to the object and based on features re-projection and filtering such as in [15] or generate the full state autonomously using correlation. The Rosetta mission approach based on local maplets of selected features is illustrated in [18]. Small scale high-resolution elevation maps called maplets centered on specific landmarks are spread across the surface of the comet. Combining together a 3D map, an albedo map, and a photometric model it is then possible to predict the landmark appearance at varying illumination conditions. This is used to render the maplet and perform correlation. The methods developed for the Osiris-Rex mission instead is based on natural features rendering and correlation, as illustrated in [19]. An approach used in the Hayabusa 2 mission that makes uses of artificial landmarks left on the surface of the body is instead detailed in [20].

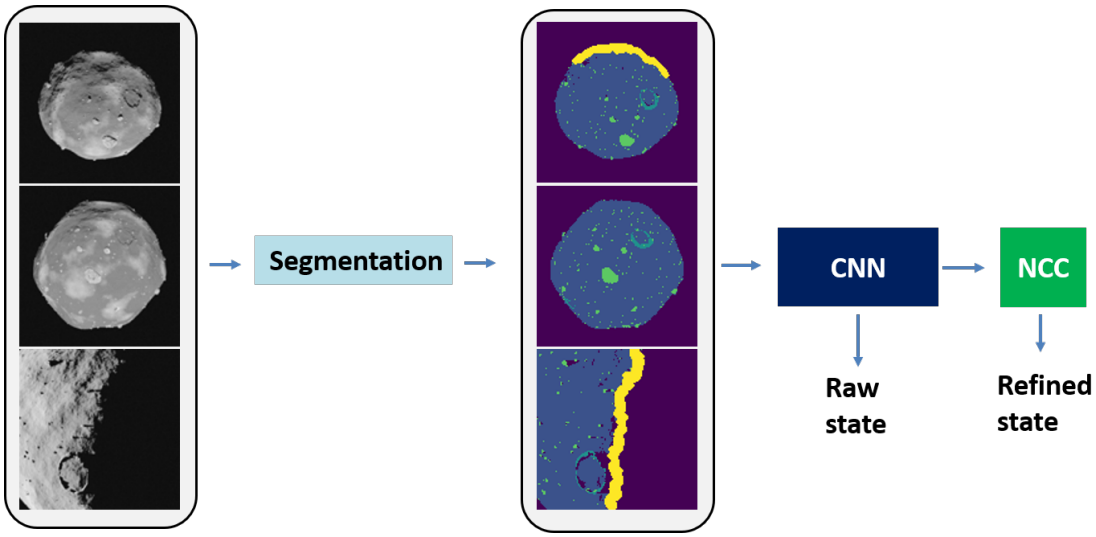

Figure 2: Macro-steps of the proposed method for small-body navigation with segmentation maps.

Most of these techniques have in common the usage of a model to be rendered (on-board or on-ground) for correlation with real images. The more complex is the model and the procedure to use, the computationally more expensive is to run the technique on-board. For example, the optical navigation performed in Rosetta required in its initial operational steps a semi-supervised approach with human-in-the-loop that was performed on-ground in order to achieve very high accuracy [18]. Correlation, as well as pixel-intensity based metrics, are highly influenced by the illumination condition, which shall accurately be modeled in order to have robust navigation solutions. Moreover, to be used on-board with a certain degree of confidence, it may be necessary to initialize them in a

basin of convergence as close as possible to the true solution. This is helpful to reduce the search space size and to make them more computationally efficient.

Recent trends in optical navigation explore the use of artificial intelligence and in particular deep learning to tackle these challenges and bypass traditional image processing techniques, providing directly the estimated state of the spacecraft as output. For example in [21] a single layer feed-forward network, also known as Extreme Learning Machine, is used to determine a spacecraft horizontal coordinates over a lunar terrain with an ideal nadir pointing. A subsequent work went even further by bypassing the navigation and generating control as output directly from images [22].

The main advantage of these techniques is the embedding of a complex task within the weights and bias of a network architecture. This comes at the cost of a large amount of realistic data needed for training.

This work proposes a novel method for on-board autonomous navigation based on segmentation maps. The method is enabled by Deep-Learning, in the form of a CNN, that is used to predict the class of the estimated state of the spacecraft in an asteroid fixed reference frame, followed up by a refinement step using a NCC based method, as illustrated in Figure 2. Didymos asteroid and Hartley comet are considered. Instead of working with binary or grayscale images, the method uses segmentation maps. Because pixel intensity is now associated with semantic meaning, an image can be deconstructed into its main components (background, surface, craters, boulders, terminator region) and this information is used for navigation.

The paper is organized as follows. The dataset generation and the class subdivision is illustrated in the next section. Subsequently, the key components of the navigation method are presented, followed up by a section discussing the concept of telescopic range. The assessment of the results on the 5 test scenarios considered is given in a dedicated section, followed up by some conclusion and final considerations for future works.

DATASET GENERATION

Didymos and Hartley are considered as representatives of regular and irregular small-bodies. For simplicity Dimorphos, the secondary body of the Didymos binary system [23], is not considered in this work. A camera sensor with a FOV of 10° is assumed and the space around each shape model is normalized based on the distance at which the maxim bounding box of the model touches the edges of the image. This normalization distance is referred to as D_0.

To account for a realistic proximity operation scenario, some constraints are set. The lower and upper boundaries for the range are set to 0.7 and 1.3 to encompass both scenarios in which only features are visible or when the body global outline can be fully discerned in the image. Furthermore, it is assumed that the spacecraft would most likely lie within an equatorial region with an elevation bounded between $\phi = \pm 45$. Enforcing these constraints the spherical shell structure illustrated in Figure 3 is obtained.

A total of 1176 are carved from this structure in polar coordinates. Intervals of 0.1, 15° and 15° are considered respectively for the range, equatorial and elevation angles ($R - \theta - \phi$), generating respectively 7 macro-classes for range and elevation angles and 24 macro-classes for the equatorial angle.

For each class, varying illumination conditions are considered assuming the sun direction to be lying in the equatorial plane of the spherical shell and with admissible phase angles between $\pm 90^\circ$.

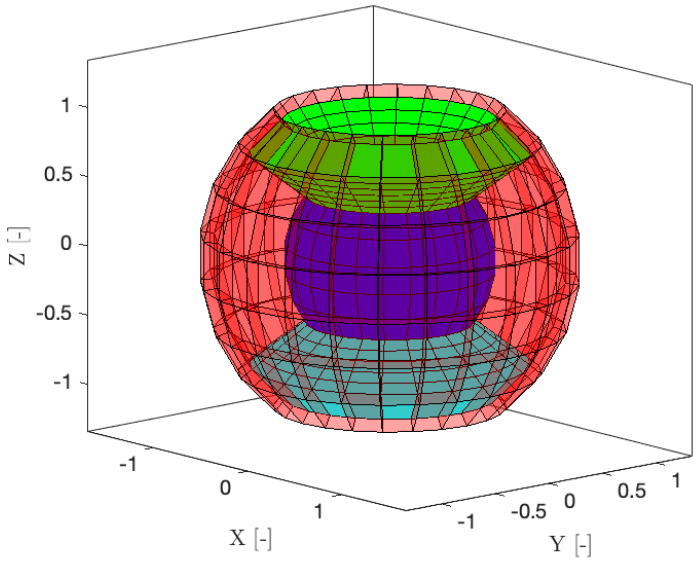

Figure 3: Structure of the 1176 classes considered distributed over the spherical shell. The red and blue ones are the innermost and outermost one from the small-body. The green and cyan ones are the top and bottom classes with the maximum elevation. In between there are the other classes that are not shown for clarity.

In such a way, impractical image acquisitions where the sun is eclipsed by the small-body are avoided. The pointing is considered ideal. A various set of images are obtained and are summarized in Table 1.

Table 1: Number of images per test case for each small-body shape

Set	Number of images	Position	Sun
Train 1	8232	Centered (Class)	Regular
Train 2	35280	Uniform (Class)	Uniform
Valid 1	1500	Uniform (All)	Uniform
Valid 2	4704	Uniform (Class)	Uniform
Test 1	5000	Uniform (All)	Uniform
Test 2	432	Trajectory	Trajectory

Train 1 set is composed of 7 image-mask pairs per class (taken at the center of the class in polar coordinates) with 7 equally spaced phase angles. Valid 1 and Test 1 are taken with randomly selected positions satisfying the spherical shell constraints and with random illumination conditions. Train 2 and Valid 2 sets are obtained by selecting 30 and 4 random positions respectively within each class with random illumination conditions. Test2 is made by positions taken from a trajectory representative of an operational scenario around Didymos.

The image-mask pairs are generated synthetically using Blender*, an open-source, easy to use, flexible software supporting python scripting. Both are 128×128 pixel wide, the images are in grayscale while the segmentation masks assume 5 different discrete levels (0,64,128,191, and 255)

*https://www.blender.org/

corresponding respectively to the background, body surface, craters, boulders, and terminator region. The interested reader can refer to [24] for the methodology used to generate the masks.

In this work, mostly segmentation maps are used. These detail the positioning of morphological structures in the image, as it is possible to see in Figure 1. Several approaches can be used to perform such segmentation. In this work, the one presented in [24] is considered, where a U-Net architecture is designed and tested over several small-bodies. These maps are envisioned to be used for hazard avoidance and landing, for autonomous scientific acquisition and planning, and, as in the case of this work, for navigation.

METHODOLOGY

An on-board method is proposed that is capable of estimating a spacecraft position with respect to an asteroid fixed reference frame exploiting segmentation maps of a previously well-known small-body. The method is divided into two parts, as illustrated in Figure 2. In the first one, a CNN is used to provide a rough estimate of the position. In the second one, a NCC based algorithm is used to refine the CNN estimate once a basin of convergence has been identified.

The space around the small-body is divided into 1176 classes and the CNN is employed as a classifier with the main purpose to greatly reduce the parameter's search space for the NCC to be applied. Classification was preferred with respect to regression for the possibility to have the top-predicted classes accompanied by the lower scoring ones to build up class distributions. These can be used to illustrate how much the predictions are spread across different classes to build up a confidence region that accompanies the predicted solution. In this method, the CNN acts both as a navigation method robust to various conditions and a proxy of an on-board rendering tool, as in its 725880 parameters is embedded the appearance of more than 43512 128×128 segmentation maps and their class distributions.

Previous works also preferred classification over regression, as for the cases of [25], in which a CNN is used to estimate the position of a spacecraft with respect to the Moon using images distributed across classes over a digital terrain map. Another interesting application comes from the work in [26], in which image geolocation is performed with cells spread across the planet's surface whose dimensions are a function of the density of the images. In this way, not only it is possible to see the most likely locations, but also the ones that correlate most with it due to geographical similarity.

However, the classification framework has one major drawback: by design, it cannot achieve high accuracy. Even when the correct class is identified, the maximum error is only guaranteed to be smaller than half the class dimensions. This poses an important limitation, that is overcome easily with the adoption of the NCC based method.

The advantage of using segmentation maps in place of grayscale images is due to the fact that pixel-intensity variations due to illumination conditions do not play a direct role and that simple, low-resolution models can be considered for online rendering.

Convolutional Neural Network

A CNN is a particular type of deep neural network commonly applied for image analysis. It has been largely used as the standard for image classification and many other tasks involving spatial input [27]. A schematic representation of a typical CNN is illustrated in Figure 4.

The architecture considered in this work has been designed in TensorFlow 2.1 [28] exploiting for the convolutions layers the architecture of the MobileNetV2 network [29], an advanced CNN architecture for image classification.

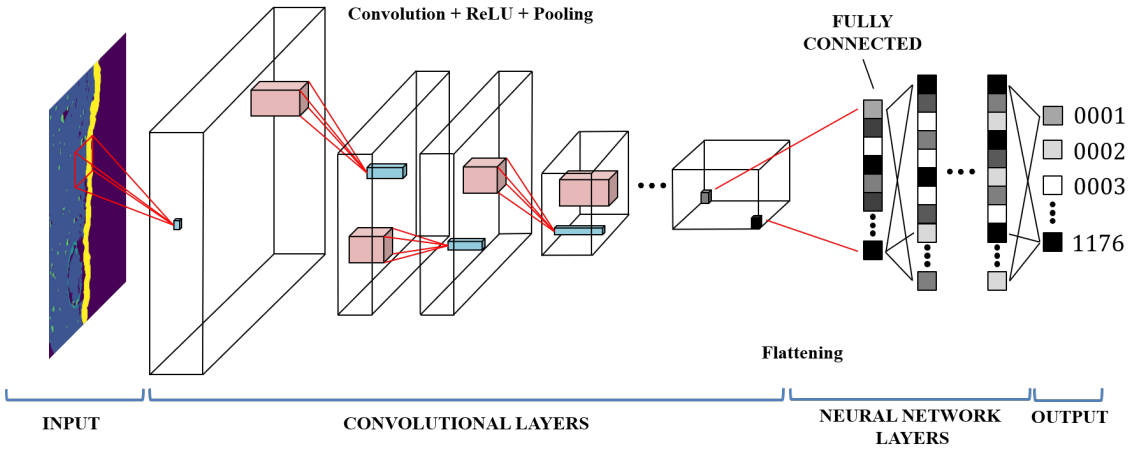

Figure 4: Schematic representation of the CNN architecture for classification. The Input is a segmentation map, the output is the associated class. The convolutional layers perform several image processing tasks to correlate spatial information, the neural network layers sort out this information to generate a class identifier.

The hyper-parameters search has been divided into two parts following an approach highlighted in [30] and adopting the philosophy described in [31]. In the first one, without any regularization in place, an optimal set in the parameter space for the learning rate, batch size, α, depth of dense layers is performed for a very short amount of time. After optimal regions of these parameters are identified, epochs are increased and regularization is introduced in the form of dropout rates and image shift. This process is repeated by refining the parameter space alongside a reduction of the number of cases and an increase in the training epochs. During the whole procedure, global metrics are considered for both training and validation accuracy as well as the convergence speed. The most promising ensembles of the network are then fine-tuned to achieve the desired performances.

Table 2: Hyper-parameters of the CNN architectures used in this work.

Parameter	**Value**
Optimizer	Adam
Loss function	Sparse Categorical Cross Entropy
α (MobileNetV2)	0.35
Dropout rate (1)	30% (Didymos), 15% (Hartley)
Dropout rate (2)	50%
Learning rate	0.0005
β_1	0.9
β_2	0.999
Pixel shift	5
Batching strategy	mini-batch
Batch size	128
Steps per epoch	339
Training epochs	100

The same architecture is used for Didymos and Hartley, with different weights and biases and different tuning of the hyper-parameters, summarized in Table 2. The CNN architecture and the

accuracy and loss during training are illustrated in the Appendix. The weights and bias used are the ones achieving maximum accuracy on the validation sets.

Normalized Cross-Correlation

In a correlation procedure, an image is taken in the environment and is confronted with a set of images either saved in a database or rendered online. The first one can be referred as the real image while the latter as the template.

Because the parameters of the templates can be controlled, they can be varied to maximize the similarity between templates and real images. To assess this similarity different metrics can be used. In this work a correlation metric derived from [32] is adopted:

$$\gamma = \frac{\sum_{ij}(R_{ij} * T_{ij})}{\sqrt{\sum_{ij} R_{ij}^2 * \sum_{ij} T_{ij}^2}} \tag{1}$$

Because in this work correlation is performed between segmentation maps, albedo difference between template and real map does not play a direct role. It could do so indirectly if the segmentation method used is not robust enough to variations of the illumination conditions. For this reason, the formula has been modified to avoid such sensitivity. For practicality, the correlation metric is computed in the spatial domain and not in the frequency one.

The correlation is expected to be handled on-board, for this reason, the model used for rendering the templates is simpler than the one used to emulate the real segmentation maps processed from images taken in the environment. In practice this is obtained by relaxing the rendering settings, changing the material properties, the mesh structure, removing the terminator layer, and also by considering only the biggest boulders, as can be seen from Figure 5.

Figure 5: Comparison between the real segmented maps (Left) and the template ones with maximum correlations at iterations 1, 2 and 3. The correlation coefficient varies from left to right from 0.58, 0.69 and 0.75 while the error from 281, 111 and 21 m respectively.

For comparison, a rendering of the real segmentation map takes roughly 15s, while the rendering of the simplified model just 0.3s considering a Cycles CPU rendering with an Intel(R) Core(TM) i7-8700 CPU @ 3.20GHz.

Other important assumptions that have been taken in this work are that the illumination conditions are given and that the correlation search space is only driven by variations in $R-\theta-\phi$ components. The pointing is assumed ideal, even though the CNN can accommodate non-ideal pointing. For a real application, these parameters shall also be considered as part of the search space.

The procedure used is the following. As a starting point, the predicted class from the CNN and all the classes immediately next to it are used to define the initial intervals of the search space. A random distribution of 100 points is then generated in polar coordinates within this region and the correlation coefficients are computed using Equation 1 between real and template maps. The point with the highest correlation is saved and a new iteration is performed by reducing the search space interval of each component to $1/3$ of the original one. The new interval is used and the procedure is repeated for a total of 3 iterations, making sure that the point with the absolute best coefficient gets to survive from one iteration to the next.

Variants of the method with different interval reduction strategies, in Cartesian coordinates and with pixel-intensity based metrics have also been tested but did not show promising results when compared to the NCC metric.

TELESCOPIC RANGE

The normalization on the space around Didymos and Hartley can be useful to generalize the results to different range-camera settings. However, the CNN cannot be applied as it is outside the boundaries imposed by the spherical shell in Figure 3. A method to extend the functioning range of the CNN is therefore introduced in this section: the telescopic range.

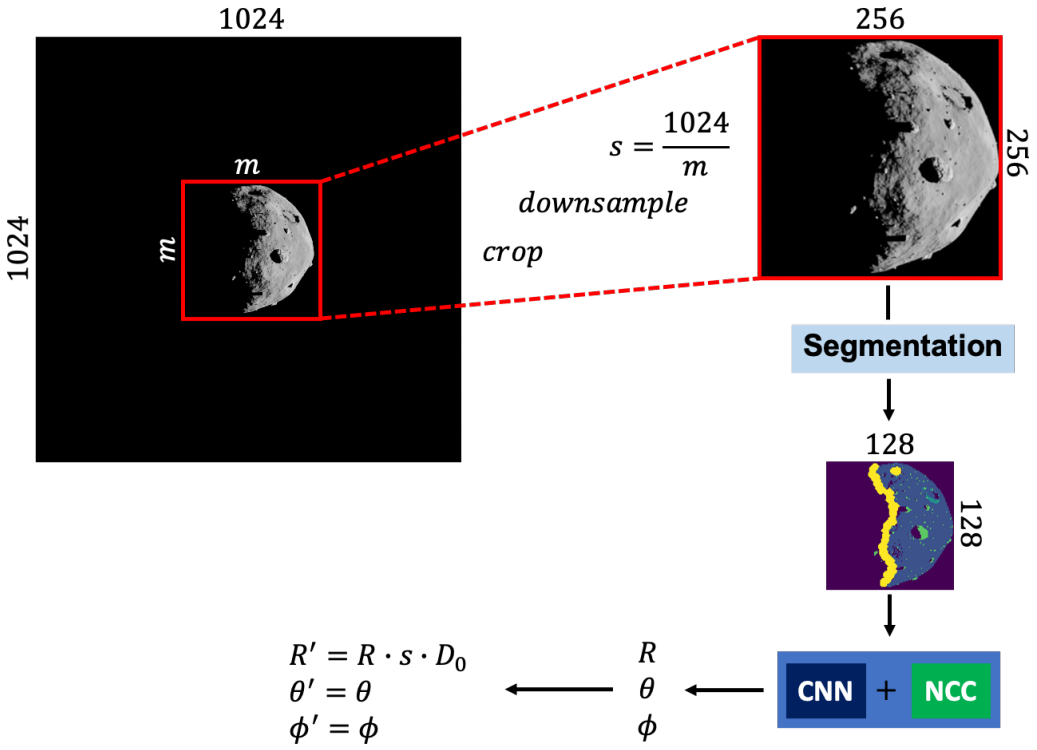

Figure 6: Example of telescopic range applied to an image in which Didymos is seen from a range much further away than the one considered for training.

Whenever an image is taken from a distance larger than $1.35D_0$, a series of simple image processing tasks is performed to transform the image appearance to one that will be suitable for the CNN.

A combination of cropping, down-sampling and re-scaling is used to do so. The main assumption is that apart from scaling, the appearance of the bodies do not differ too much due to perspective changes given by different ranges-FOV configurations.

The procedure is illustrated in Figure 6. When an image is taken, a bounding box of side m is identified around the object. If this value correspond to an object seen with $0.65 < R < 1.35$ then a simple down-sampling is performed to reduce the image size to the input needed by the segmentation method, in this case 256×256 pixels. If this is not the case, but $m > 256$, then the image is cropped around the object with the box of side m. The scale factor between the original image size and m is then recorded. The cropped snippet is now down-sampled until a size of 256×256 pixels is obtained. A segmentation method is then performed on the image to obtain a mask with dimension 128×128 pixels. The CNN and NCC techniques are then applied to the mask to predict R, θ, and ϕ. The predicted values of R is then the only one that need to be scaled back and is multiplied by s and D_0 to obtain the dimensional range R'. This $R' - \theta' - \phi'$ set is now capable to describe the spacecraft state in an asteroid fixed reference frame in world dimensions, even though the original image format was not suitable for the CNN. A case in which the telescopic range is tested is presented in the results section for an operative orbit about Didymos.

RESULTS

In this section, the method presented in this work is tested on 5 different scenarios, summarized in Table 3.

In all datasets but in D-3 the segmentation maps used are the true ones, while in D-3 the predicted ones with the U-Net described in [24] is used. The CNN performances are assessed with datasets D-1, D-2, D-3, and D-4 while the NCC with D-5.

Table 3: Summary of the datasets used in the five test cases, identified as *D-* followed by the test identification number.

Dataset	Models	Section name	Number of maps
D-1	Didymos, Hartley	Test set	5000, 5000
D-2	Didymos, Hartley	Reduced maps	15000, 15000
D-3	Didymos, Hartley	Predicted maps	5000, 5000
D-4	Didymos	Close-proximity operations	432
D-5	Didymos, Hartley	Normalized Cross-Correlation	1000, 1000

Test set

Dataset D-1 is made by 5000 segmentation maps of Didymos and Hartley from Test 1 of the dataset generation step. With this dataset, the CNN achieves an accuracy of 75.94% with Didymos and of 68.60% with Hartley. However, the standard definition of accuracy is not capable to describe fully a more complex behavior of the classifier.

In the specific classification task presented in this work, spatial proximity between classes is important but is not encoded with the class label. Classes next to each other do not represent different species of objects but do represent the same object seen from a slightly different perspective. Thus, a classification next to the correct class shall not be discarded as a completely wrong classification, because it demonstrates the capability of the CNN to provide a solution close to the correct one. In

order to take this into account, a metric that also consider proximity between classes is designed in this work: the Inter-Class Distance or ICD.

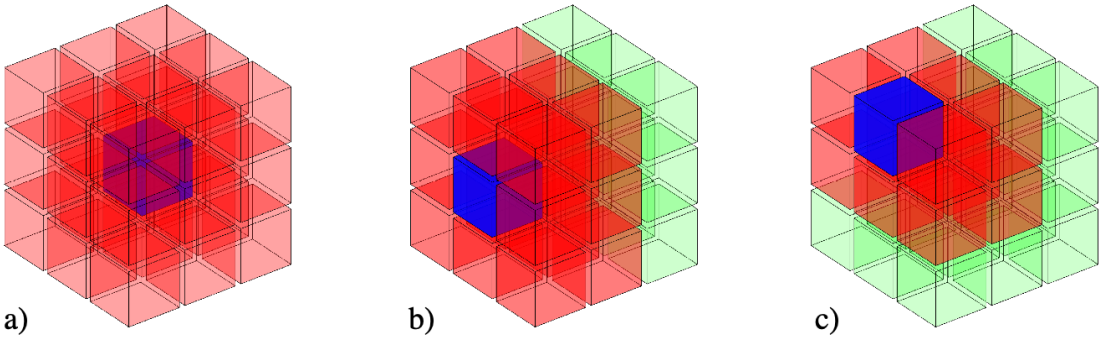

Figure 7: Various examples to illustrate the concept of ICD from the blue cube: blue (ICD0), red (ICD1), green (ICD2). a) central class surrounded by 26 ICD1. b)lateral class surrounded by 17 ICD1 and 9 ICD2. c) edge class surrounded by 11 ICD1 and 15 ICD2.

The ICD is defined as the minimum number of classes that shall be passed through that connect a given point P in a class A with another point Q in a class B. A movement from one class to another is allowed if the two shares either a face, an edge or a vertex. The concept is illustrated in Figure 7. If you imagine a Rubik's cube composed by 27 smaller cubes and consider a point in the center cube (the one in which the spherical joint is hidden), then the remaining 26 cubes can be reached with a single jump. Because of the characteristic topology of the spherical shell considered in this work, only 3 scenario are possible to start the count of the ICD from each class, as schematized in Figure 7. A matrix describing the ICD relationships of each class with the other $1176 - 1$ ones of the spherical shell is generated.

The accuracy is then defined considering the ICD metric. The $a[ICDn]$ is the accuracy over the classes with an ICD equal to n while the notation $a[ICDn+]$ and $a[ICDn-]$ are used for the cumulative accuracy taking into account all classes respectively above or within a distance of $ICDn$.

Table 4: CNN performances expressed with the ICD metric for dataset D-1.

Metric	Didymos	Hartley
$a[ICD0]$	75.94	68.60
$a[ICD1]$	23.96	31.28
$a[ICD2+]$	0.08	0.12

The performances of the CNN with this new metric are summarized in Table 4. It is possible to see that roughly 99.92% and 99.88% of the classifications happen in the correct class or within 1 class from the correct one. Very few cases (4 and 6 respectively) happen at classes much further away from the correct ones. Moreover, illumination conditions seem to be relevant for Didymos prediction with an $ICD1$, while the same cannot be said for Hartley, as can be seen in Figure 8.

This is an important result since it demonstrates that the CNN is definitely capable to reduce the search space of the estimated position by grouping almost all cases in or next to the correct class and a considerable number of cases in the correct one.

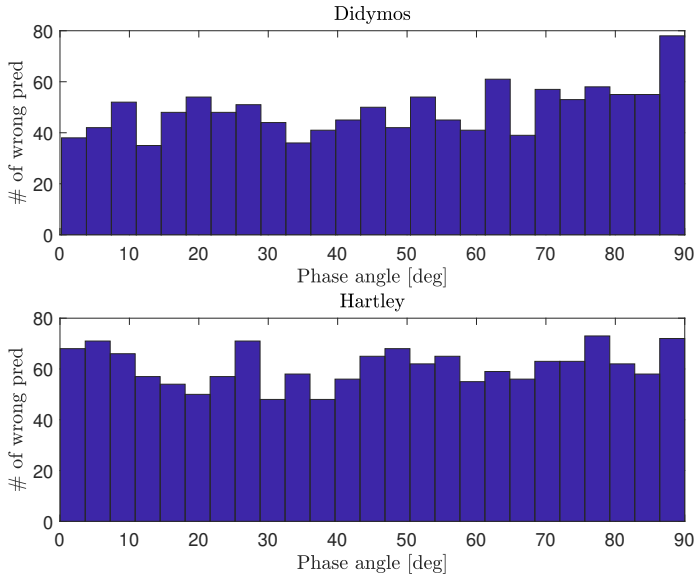

Figure 8: Histograms of the number of prediction with $ICD1$ for Didymos and Hartley as function of the phase angle.

Reduced maps

Dataset D-2 is made by 3 subsets of 5000 segmentation maps each of Didymos and Hartley. The point of view and illumination conditions are the same as D-1, but the 3 subsets are obtained by removing craters and boulders (No features), by removing only craters, and by removing only boulders. In this way, it is possible to isolate the impact that single layers have on the classification. Table 5 and Table 6 summarize the cumulative accuracies for Didymos and Hartley.

Table 5: CNN performances expressed with the ICD metric for dataset D-2 on Didymos.

Metric	No features	No craters	No boulders
$a[ICD0-]$	2.2	66.2	5.1
$a[ICD1-]$	11.8	97.5	21.0
$a[ICD2-]$	20.2	97.8	29.6
$a[ICD3-]$	37.5	98.4	45.7
$a[ICD4-]$	57.3	98.7	60.9
$a[ICD5-]$	72.8	99.0	74.6
$a[ICD6-]$	88.1	99.7	90.5
$a[ICD7-]$	100.0	100	100

From this test is possible to understand, after training, which layers play a role and which not. Didymos classifications are heavily influenced by craters while boulders constitute the most influential layer. On the other hand, Hartley classifications are not so heavily affected by boulders and are completely indifferent to craters. In this case, the backbone of the classification seems to rely on the body outline.

When boulders and craters are not considered and the camera is so close to the body that the background layer does not appear, the only remaining layers, surface, and terminator may not be enough to determine the position. This seems to be more relevant to Didymos than to Hartley.

Table 6: CNN performances expressed with the ICD metric for dataset D-2 on Hartley.

Metric	No features	No craters	No boulders
$a[ICD0-]$	60.8	69.1	61.8
$a[ICD1-]$	99.3	99.8	99.4
$a[ICD2-]$	99.5	99.9	99.6
$a[ICD3-]$	99.6	99.9	99.6
$a[ICD4-]$	99.6	99.9	99.7
$a[ICD5-]$	99.6	99.9	99.7
$a[ICD6-]$	99.6	99.9	99.7
$a[ICD7-]$	100.0	100	100

Predicted masks

Dataset D-3 is generated by taking the images associated with the masks of D-1 and evaluate them over the U-Net described in [24]. The purpose of this test is to demonstrate that the CNN can be trained with the true masks and that it can then be deployed for a real application and still work fine with the predicted masks.

Table 7: CNN performances expressed with the ICD metric for dataset D-3 on Didymos.

Metric	Didymos	Hartley	Metric	Didymos	Hartley
$a[ICD0]$	52.5	48.1	$a[ICD0-]$	52.5	48.1
$a[ICD1]$	38.3	48.9	$a[ICD1-]$	90.8	97.1
$a[ICD2]$	0.8	1.4	$a[ICD2-]$	91.5	98.5
$a[ICD3]$	1.4	0.3	$a[ICD3-]$	93.0	98.8
$a[ICD4]$	1.4	0.2	$a[ICD4-]$	94.3	98.9
$a[ICD5]$	1.6	0	$a[ICD5-]$	95.9	99.0
$a[ICD6]$	2.1	0.1	$a[ICD6-]$	98.0	99.1
$a[ICD7]$	2.	0.9	$a[ICD7-]$	100	100

As it is possible to see from Table 7, the CNN suffered a drop in performances when predicted masks are used, but overall high values of $a[ICD1-]$ have been retained. An interesting possibility that is not investigated in this work is to re-train the CNN with the predicted masks to overcome such drop.

Close-Proximity operations application

Dataset D-4 is made by 432 images of Didymos taken from the trajectory illustrated in Figure 9 with a FOV of $16 \times 16^{\circ}$ from a 1024×1024 sensor. The orbit is a preliminary trajectory that has been considered for the Milani mission [7] and is used in this work to provide an example of a typical geometry that can be encountered in the proximity of a small-body.

The trajectory displayed in Figure 9 is composed of an inner arc of 1 day immediately followed by an outer arc of 2 days. Didymos rotational period is assumed to be 2.22 hours. Images are taken every 10 minutes with ideal pointing in a 3-day loop, for a total of 432 image-mask pairs. The results of the CNN are represented in polar coordinates in Figure 10, while the estimated positions are superimposed on the true trajectory in Figure 9.

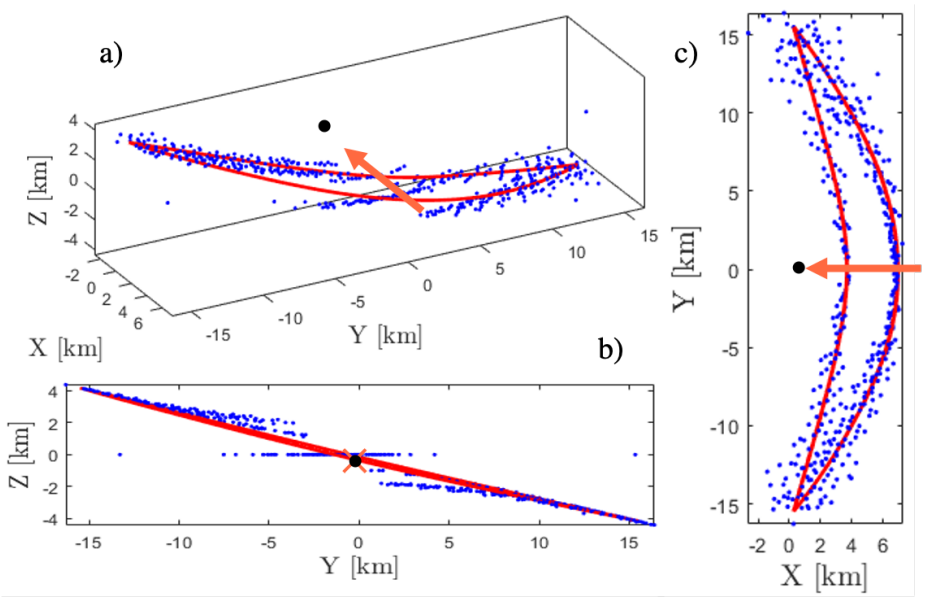

Figure 9: Close-proximity orbit example around Didymos (red) and estimated positions with the CNN (blue). The trajectory is illustrated in the Asteroid-Sun fixed reference frame, with the X axis pointing towards the Sun. a),b) and c) illustrate the trajectory seen from a 3D perspective, YZ plane and XY plane respectively.

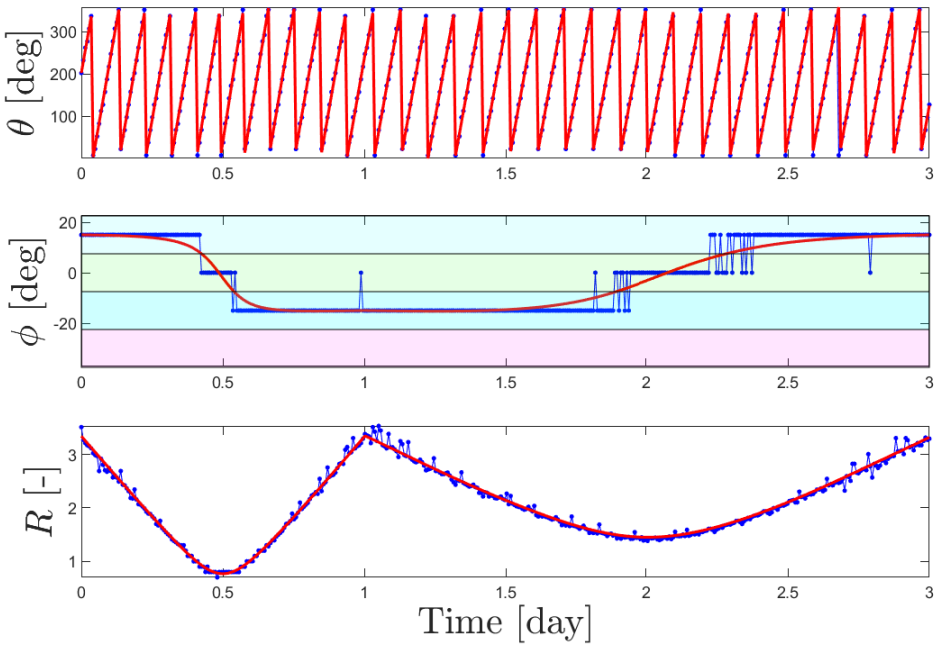

Figure 10: Predicted (blue) and true (red) equatorial, elevation and range values as function of time for the close-proximity trajectory.

As it possible to see in Figure 10, a relevant portion of the trajectory happens outside the $[0.65 - 1.35]D_0$ interval considered during training. Whenever the range exceed the upper limit of this interval, the telescopic range concept illustrated before is used. The CNN worked out well for what concern prediction of the equatorial angle θ and range R. The CNN also worked as expected for the prediction of the ϕ angle. However, the intervals used for the latter do not fit well enough the trajectory values, that develops around $\phi = \pm 15°$.

Normalized Cross-Correlation

Dataset D-5 is a subset of D-1 made by 1000 random masks of Didymos and Hartley to which the NCC method described in the methodology section is applied. The mean positioning error achieved with Didymos and Hartley is respectively 63.49 m and 301.06 m, after an outlier rejection with success rates of 95.30% and 92.20%. In Figure 11 is possible to see the relative percentage error with respect to the true range ϵ_r as a function of the predicted range in adimensional space for both bodies. The mean relative percentage error is respectively 1.37% and 2.15% for Didymos and Hartley. The normalized range of the two bodies is respectively $D_0 = 4.81$ km and $D_0 = 14.41$ km.

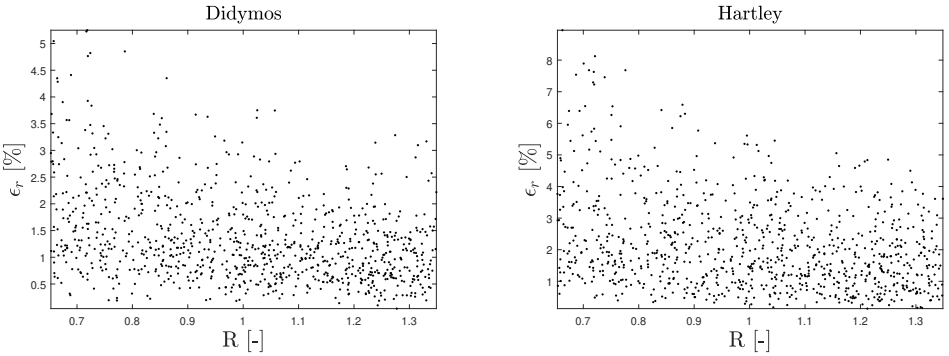

Figure 11: Relative percentage error as function of the range R from the bodies. The normalized range D_0 corresponds to $R = 1$.

CONCLUSION AND FUTURE WORKS

This work proposed a novel method for on-board autonomous navigation based on segmentation maps. The method is enabled by a combination of deep-learning and correlation methods. A CNN is used to predict the class of the estimated state of the spacecraft in an asteroid fixed reference frame, this estimation reduces the search space for a NCC based method to be applied for a more accurate navigation solution. Didymos and Hartley have been considered and the performance of the methodology has been tested in 5 different scenarios.

Practical assumptions for the CNN are to have available on-ground a segmented model of the asteroid for the generation of the synthetic masks and that the trajectory develops below certain elevations, constrained by the spherical shell considered. The first can be addressed as a traditional shape model in which the most prominent boulders and craters are highlighted. This further complexity on the model on-ground is justified by the fact that segmentation maps used on-board are not subject to pixel-intensity variations and that it would be possible to use simpler, low-resolution models for on-board rendering. The practical assumptions for the NCC based method are that the

pointing is ideal and illumination conditions are assumed to be known (either estimated from Sun sensors or image processing) and that only the camera position is changed during online rendering.

The CNN demonstrated to be capable to generate robust predictions, but most importantly has demonstrated to be capable of grouping these in the space immediately around the correct classes. Depending on the small-body considered, the CNN was shaped through training on the importance of the layers of the segmentation maps to be considered to produce a position estimate. It was found that surface features such as boulders play a key role for Didymos, while the same cannot be said about the network trained over Hartley, which has shown a classification based on the other layers. This different behavior seems to be linked with the regularity of the shape models. It was assessed that a moderate drop in performance can be expected when predicted masks are considered instead of true masks. When put to the test with the geometry of a real close-proximity trajectory, the CNN demonstrated to have good predicting performances even when images had been modified to be translated within the training region. Important limitations have also been observed in this scenario due to the small set of ϕ considered. Finally, the NCC method has demonstrated that when coupled with the CNN one can greatly increase the positioning accuracy of the latter and provide a robust estimate.

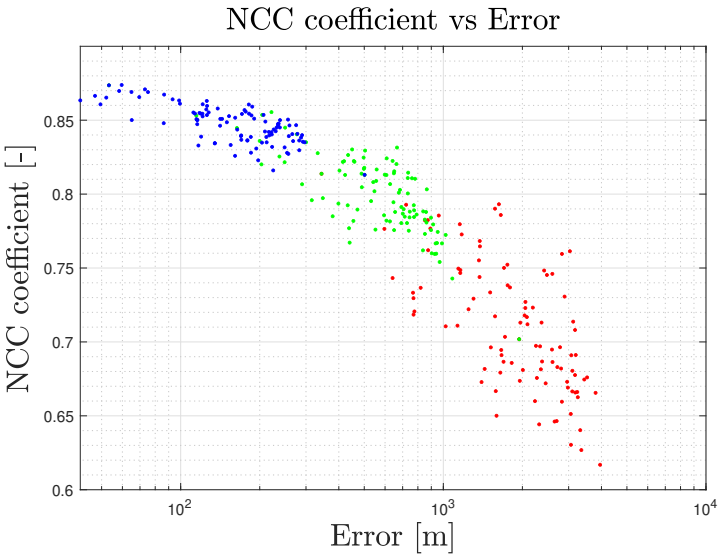

Figure 12: Example of the NCC coefficient as function of the error from iteration 1 (red), 2 (green) and 3 (blue).

The methodology proposed does not distinguish directly between features in the classical sense (such as boulders and craters) and outline. Each layer of the segmentation map is instead considered as a type of feature. In this way, the classical distinction between features and outline based methods is not respected and the method can be applied on a variety of ranges from the body, independently from the image apparent size in the FOV, without a change in the algorithm nor in the settings.

One issue that has been encountered with the adopted NCC method is related to the fact that the minimum error is not guaranteed to coincide with the maximum NCC coefficient, as it is possible to see from one case in Figure 12. This aspect is crucial in defining a correct metric for an optimization procedure that minimizes the error and future work will be directed in this area.

Training could be improved from considerations of the ICD and by reduction of the number of samples per class. It has been observed that in order to have high-values of $a[ICD0]$ on the validation and test sets it was important to increase the distances between classes. One way to do this is by increasing the sample density in each class as well as network complexity. However, more efficient alternatives could be to define specific loss functions that encapsulate the ICD or to deploy annealed classes with fuzzy borders.

Future work will be focused on a re-design of the NCC based method, with a better design of features extracting and matching techniques and potentially also other deep learning architectures in its place. Also different and more irregular bodies shall be considered for testing, to better understand if certain trends generalize to different shapes and features.

ACKNOWLEDGMENT

The authors would like to thank Fabio Ferrari for providing the closed shape trajectory for the telescopic range test. A special thanks also to Michele Maestrini for the invaluable and interesting discussions and Carmine Buonagura for the procedural generation tool whose initial steps were used to generate the changes on Didymos and Hartley shape models. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 813644.

REFERENCES

[1] M. B. Quadrelli, L. J. Wood, J. E. Riedel, M. C. McHenry, M. Aung, L. A. Cangahuala, R. A. Volpe, P. M. Beauchamp, and J. A. Cutts, "Guidance, Navigation, and Control Technology Assessment for Future Planetary Science Missions," *Journal of Guidance, Control, and Dynamics*, Vol. 38, July 2015, pp. 1165–1186.

[2] S. Casini, I. Fodde, S. Engelen, B. Monna, A. Cervone, and E. Gill, "Towards the Use of Commercial-off-the-Shelf Small-Satellite Components for Deep-Space CubeSats: a Feasibility and Performance Analysis," *SmallSat 2020-34th Small Satellite Conference*, 2020.

[3] S. Casini, I. Fodde, B. Monna, A. Cervone, and E. Gill, "Novel 3U Stand-Alone CubeSat Architecture for Autonomous Near Earth Asteroid Fly-By," *Aerospace*, Vol. 8, Dec. 2020, p. 9, 10.3390/aerospace8010009.

[4] R. Walker, D. Koschny, C. Bramanti, I. Carnelli, E. C. S. Team, *et al.*, "Miniaturised Asteroid Remote Geophysical Observer (M-ARGO): a stand-alone deep space CubeSat system for low-cost science and exploration missions," *6th Interplanetary CubeSat Workshop, Cambridge, UK*, Vol. 30, 2017.

[5] E. A. Woeppel, M. Ott, J. Balsamo, J. A. Styborski, C. A. Roche, M. J. Scorza, C. Doherty, K. Fisher, A. J. Trovato, C. Volk, *et al.*, "The Near Earth Object Scout Spacecraft: A low-cost approach to in-situ characterization of the NEO population," 2014.

[6] P. Tortora and V. Di Tana, "LICIACube, the Italian witness of DART impact on Didymos," *2019 IEEE 5th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, IEEE, 2019, pp. 314–317.

[7] F. Ferrari, V. Franzese, M. Pugliatti, C. Giordano, and F. Topputo, "Preliminary mission profile of Hera's Milani CubeSat," *Advances in Space Research*, Jan. 2021, 10.1016/j.asr.2020.12.034.

[8] H. Goldberg, Ö. Karatekin, B. Ritter, A. Herique, P. Tortora, C. Prioroc, B. G. Gutierrez, P. Martino, I. Carnelli, and B. Garcia, "The Juventas CubeSat in Support of ESA's Hera Mission to the Asteroid Didymos," *Small Satellite Conference, Logan, Utah*, 2019.

[9] J. Gil-Fernandez and G. Ortega-Hernando, "Autonomous vision-based navigation for proximity operations around binary asteroids," *CEAS Space Journal*, Vol. 10, Feb. 2018, pp. 287–294, 10.1007/s12567-018-0197-5.

[10] S. Bhaskaran, J. E. Riedel, and S. P. Synnott, "Autonomous nucleus tracking for comet/asteroid encounters: the Stardust example," *1998 IEEE Aerospace Conference Proceedings (Cat. No.98TH8339)*, Vol. 2, 1998, pp. 353–365 vol.2, 10.1109/AERO.1998.687921.

[11] L. A. . A. B. Wright, C., "Optical navigation algorithm performance," 2018. 1st Annual RPI Workshop on Image-Based Modeling and Navigation for Space Applications, Troy, NY.

[12] J. A. Christian, "Optical Navigation Using Planet's Centroid and Apparent Diameter in Image," *Journal of Guidance, Control, and Dynamics*, Vol. 38, Feb. 2015, pp. 192–204, 10.2514/1.g000872.

[13] J. A. Christian and S. B. Robinson, "Noniterative Horizon-Based Optical Navigation by Cholesky Factorization," *Journal of Guidance, Control, and Dynamics*, Vol. 39, Dec. 2016, pp. 2757–2765, 10.2514/1.g000539.

[14] A. J. Liounis, "Limb Based Optical Navigation for Irregular Bodies," 2018. 1st Annual RPI Workshop on Image-Based Modeling and Navigation for Space Applications, Troy, NY.

[15] A. Pellacani, M. Graziano, M. Fittock, J. Gil, and I. Carnelli, "HERA vision based GNC and autonomy," 2019, 10.13009/EUCASS2019-39.

[16] W. Qian, Z. Wei, X. De, and M. Xiaoyan, "Model-Based Line-of-Sight Detection of an Irregular Celestial Body for Autonomous Optical Navigation," July 2015.

[17] D. A. Baker and J. W. McMahon, "Limb-based shape modeling and localization for autonomous navigation around small bodies," *2020 Astrodynamic Specialist Conference, Lake Tahoe*, AAS/AIAA, 2020.

[18] R. P. d. Santayana and M. Lauer, "Optical measurements for rosetta navigation near the comet," *Proceedings of the 25th International Symposium on Space Flight Dynamics (ISSFD), Munich*, 2015.

[19] C. Mario, C. Debrunner, *et al.*, "Robustness and performance impacts of optical-based feature tracking to OSIRIS-REx asteroid sample collection mission," *39th Annual AAS Guidance and Control Conference*, 2015.

[20] N. Ogawa, F. Terui, S. Yasuda, K. Matsushima, T. Masuda, J. Sano, H. Hihara, T. Matsuhisa, S. Danno, M. Yamada, Y. Mimasu, K. Yoshikawa, G. Ono, Y. Yokota, T. Saiki, and Y. Tsuda, "Image-based Autonomous Navigation of Hayabusa2 using Artificial Landmarks: Design and In-Flight Results in Landing Operations on Asteroid Ryugu," *AIAA Scitech 2020 Forum*, American Institute of Aeronautics and Astronautics, Jan. 2020, 10.2514/6.2020-0225.

[21] R. Furfaro and A. Law, "Relative optical navigation around small bodies via extreme learning machines," *Astrodynamics 2015*, Vol. 156, Univelt Inc., 2016, pp. 1959–1978. AAS/AIAA Astrodynamics Specialist Conference, ASC 2015 ; Conference date: 09-08-2015 Through 13-08-2015.

[22] R. Furfaro, I. Bloise, M. Orlandelli, P. Di Lizia, F. Topputo, R. Linares, *et al.*, "Deep learning for autonomous lunar landing," *2018 AAS/AIAA Astrodynamics Specialist Conference*, 2018, pp. 1–22.

[23] S. Naidu, L. Benner, M. Brozovic, M. Nolan, S. Ostro, J. Margot, J. Giorgini, T. Hirabayashi, D. Scheeres, P. Pravec, P. Scheirich, C. Magri, and J. Jao, "Radar observations and a physical model of binary near-Earth asteroid 65803 Didymos, target of the DART mission," *Icarus*, Vol. 348, Sept. 2020, p. 113777, 10.1016/j.icarus.2020.113777.

[24] M. Pugliatti, M. Maestrini, P. d. Lizia, and F. Topputo, "On-board small-body semantic segmentation based on morphological features with U-Net," *31st Space Flight Mechanics Meeting, Charlotte*, AAS/AIAA, 2021.

[25] T. Campbell, R. Furfaro, R. Linares, and D. Gaylor, "A deep learning approach for optical autonomous planetary relative terrain navigation," *Spaceflight Mechanics 2017*, Vol. 160, Univelt Inc., 2017, pp. 3293–3302. 27th AAS/AIAA Space Flight Mechanics Meeting, 2017 ; Conference date: 05-02-2017 Through 09-02-2017.

[26] T. Weyand, I. Kostrikov, and J. Philbin, "PlaNet - Photo Geolocation with Convolutional Neural Networks," 2016, pp. 37–55, 10.1007/978-3-319-46484-83.

[27] A. Krizhevsky, S. Ilya, and H. G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[28] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. Software available from tensorflow.org.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, https://doi.org/10.1109/CVPR.2018.00474.

[30] I. Radosavovic, R. P. Kosaraju, R. Girshick, K. He, and P. Dollar, "Designing Network Design Spaces," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, June 2020, 10.1109/cvpr42600.2020.01044.

[31] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay," 2018.

[32] J. Lewis, "Fast normalized cross-correlation, 1995," *Vision Interface*, Vol. 2010, 2010, pp. 120–123.

APPENDIX: CNN ARCHITECTURE AND TRAINING

Table 8: Architecture of the CNN considered in this work. The total number of parameters is 725880, of which 14080 are not trainable.

Layer (type)	Output Shape	Param #
input (InputLayer)	(None, 128, 128, 1)	0
input_transform (Functional)	(None, 128, 128, 3)	0
MobileNetV2_0.35_128 (Functional)	(None, 4, 4, 1280)	410208
GAP (Global average)	(None, 1280)	0
dropout_1 (Dropout)	(None, 1280)	0
dense_1 (Dense)	(None, 128)	163968
dropout_2 (Dropout)	(None, 1280)	0
dense_2 (Dense)	(None, 1176)	151704

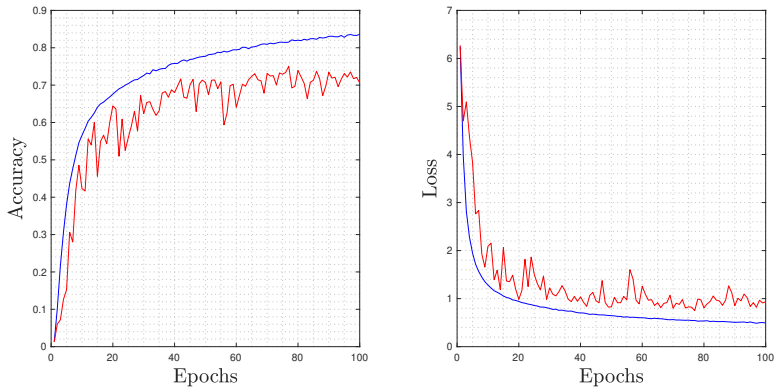

Figure 13: Accuracy (left) and loss (right) on the train (blue) and validation (red) sets of the CNN for Didymos.

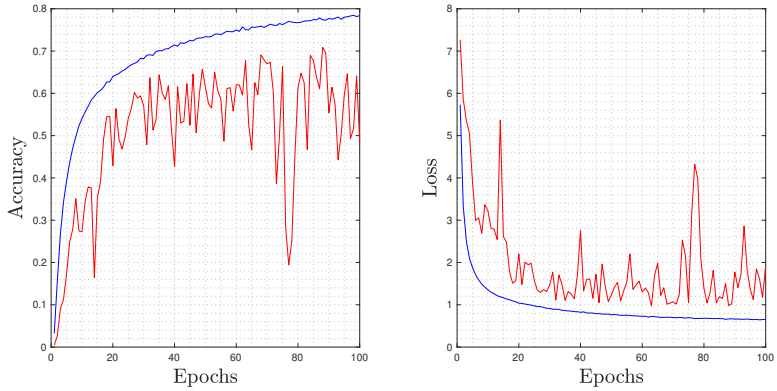

Figure 14: Accuracy (left) and loss (right) on the train (blue) and validation (red) sets of the CNN for Hartley.