



POLITECNICO
MILANO 1863

RE.PUBLIC@POLIMI

Research Publications at Politecnico di Milano

Post-Print

This is the accepted version of:

M. Piccinin, P. Lunghi, M. Lavagna
Deep Reinforcement Learning-Based Policy for Autonomous Imaging Planning of Small Celestial Bodies Mapping
Aerospace Science and Technology, Vol. 120, 2022, 107224 (12 pages)
doi:10.1016/j.ast.2021.107224

The final publication is available at <https://doi.org/10.1016/j.ast.2021.107224>

Access to the published version may require subscription.

When citing this work, cite the original published paper.

© 2022. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Permanent link to this version

<http://hdl.handle.net/11311/1191060>

Deep Reinforcement Learning-based Policy for Autonomous Imaging Planning of Small Celestial Bodies Mapping

Margherita Piccinin^a, Paolo Lunghi^b, Michèle Lavagna^c

^a*Ph.D. Student, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy.*

^b*Assistant Professor, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy.*

^c*Full Professor, Politecnico di Milano, Via La Masa 34, 20156, Milano, Italy.*

Abstract

This paper deals with the problem of mapping unknown small celestial bodies while autonomously navigating in their proximity with an optical camera. A Deep Reinforcement Learning (DRL) based planning policy is here proposed to increase the surface mapping efficiency with a smart autonomous selection of the images acquisition epochs. Two techniques are compared, Neural Fitted Q (NFQ) and Deep Q Network (DQN), and the trained policies are tested against benchmark policies over a wide range of different possible scenarios. Then, the compatibility with an on-board application is successfully verified, investigating the policy performance against navigation uncertainties.

Keywords: Small bodies shape reconstruction, Autonomous exploration, Deep learning for space applications

Email addresses: margherita.piccinin@polimi.it (Margherita Piccinin), paolo.lunghi@polimi.it (Paolo Lunghi), michelle.lavagna@polimi.it (Michèle Lavagna)

1. Introduction

Intelligent mapping is a crucial but challenging capability for small celestial bodies exploration. In fact, in proximity of small celestial bodies, the environment is extremely harsh and unknown: mission's operations related to the *mapping process* entail articulated phases, from far approach to close surveys, to gradually characterizing the body with several *mapping stages*, employing different instruments and techniques. In particular, body imaging is fundamental for the body shape reconstruction, that is performed entirely on-ground with stereophotoclinometry (SPC) [1] or stereophotogrammetry (SPG) [2] techniques. The mapping process for shape reconstruction of an unknown body requires several iterations: the shape model is refined during the subsequent observations of the body, until a high resolution model is obtained. Such process entails the collection of a large amount of images, to be sent to ground and elaborated together with navigation data in an iterative manner.

In order to achieve a good surface mapping, granting the maximum coverage and the adequate viewing and illumination conditions, trajectory design is the first necessary step. In proximity of small bodies, the gravitational field can be highly irregular and perturbations like Solar Radiation Pressure, gravitational perturbation due to the Sun and comet outgassing may play a dominant role. In such highly perturbed environments, the design of an orbit suitable for carrying out the mapping process entails many challenging aspects, related also to operational constraints and orbit maintenance, with several possible existing strategies. In the cases of binary asteroids systems, analyses dedicated to the orbit's stability are made in the three

26 body problem accounting for the irregular shape of the body[3]. Families
27 of stable orbits can be found in cases when the Solar Radiation Pressure is
28 the dominant perturbation; these orbits do not require active control and
29 therefore are inexpensive. In particular, three families of orbits are the most
30 studied in literature: ecliptic, terminator and quasi-terminator orbits [4],
31 [5], [6]. Terminator orbits lie in the plane perpendicular to the Sun and
32 are highly stable. The main drawback of this solution is that the angle be-
33 tween spacecraft and Sun is always 90° , limiting the imaging opportunities.
34 Quasi-terminator orbits are particularly good for global mapping campaigns
35 because they are stable and also offer a good variation of Sun-relative ge-
36 ometries. Nevertheless, their applicability is in practice limited, depending
37 on mission time scales, length scales and minimum allowable orbit radius.
38 Another approach is to find surrounding frozen orbits that also satisfy the
39 repeating ground track condition, which is a feature particularly useful for
40 the surface mapping [7]. Other strategies are based on actively controlled
41 trajectories, including but not limited to the heliostationary hovering [8],
42 body-fixed hovering flight [9], as well as flybys, conic-like trajectories or ping-
43 pong orbits [10], [11]. In fact, when the body mass is small such strategies
44 can still be actuated with reasonable costs and offer the possibility to easily
45 obtain the desired Sun-spacecraft-body relative geometry. Drawbacks are
46 that fuel cost may become important if the strategy is extended for a long
47 time and that maneuvers require ground supervision. A completely different
48 approach is hopping exploration over the surface [12]. As a consequence of
49 the rich and challenging dynamical environment, mapping trajectories are
50 strictly mission-dependent and related to a complex and tailored design per-

51 formed on-ground, including the planning of the orbital operations and the
52 scheduling of data acquisition for mapping.

53 While orbit selection is the first step for collecting adequate data, mapping
54 is tightly coupled with navigation and planning of the exploration [1]. Today,
55 ground support is necessary for navigation and mapping tasks, and a large
56 human effort is required even 24/7 for proximity operations supervision and
57 planning [13].

58 The aim of this work is to make a step forward in the direction of au-
59 tonomy. Autonomous explorations consists in intelligently acting in the un-
60 known environment where the agent is moving. In terrestrial robotics, tech-
61 niques for autonomous exploration have been developed for real applications:
62 in particular, in Active SLAM (Simultaneous Localization And Mapping) the
63 robot autonomously localizes itself, maps the environment and plans explo-
64 ration, tasks that are tightly coupled [14],[15],[16]. Active SLAM has also
65 been examined as being an instance of a Partially Observable Markov Deci-
66 sion Process (POMDP)[17]. In this context, a good - even if not optimal -
67 policy can be found by means of Deep Reinforcement Learning (DRL) algo-
68 rithms: the use of a DRL technique allows finding a good solution policy of
69 an otherwise computationally intractable problem [18]. Such algorithms ex-
70 hibit well proven generalizing capabilities, that are fundamental to design a
71 flexible policy capable of exploring environments with different and unknown
72 characteristics [19], and of handling problems with partial observability [20].
73 In particular, POMDPs can be tackled with DRL techniques such as Neural
74 Fitted Q (NFQ) [21] and Deep Q Network (DQN) [22].

75 In the space field, autonomous exploration has never been accomplished.

76 For systematic asteroid exploration, light and robust algorithms are required,
77 capable to promptly react to unexpected conditions, reducing risks for such
78 a delicate phase [23]. In [24], supervised machine learning is applied to
79 estimate parameters for optimal asteroid transfer trajectories. Recently, the
80 general problem of exploration has been framed as a POMDP, in analogy
81 to terrestrial active SLAM. In [25] the POMDP is reduced to completely
82 observable for designing an orbit selection policy. In [26] the POMDP is
83 tackled using DRL, overcoming the simplifications in [25] and proposing a
84 direct maneuvering policy, which can be risky for an on-board integration.

85 The focus of this work is on on-board autonomous decision-making dur-
86 ing the mapping process for shape reconstruction. Autonomous operations
87 would reduce the burden of routine navigational support and communication
88 requirements on network services, thus decreasing the mission cost. Auton-
89 omy is desirable also to maximize the mission science return (high value
90 data), enabling opportunistic science and real-time re-planning, otherwise
91 impossible because of communications delays. The margin for improvement
92 for autonomy is in first place related to the autonomous scheduling of im-
93 ages acquisition epochs. Nowadays, *images acquisition* policy is established
94 on ground during operations [13]. Hundreds of thousands of images are col-
95 lected during a mapping process. This work proposes a general approach,
96 to be applied notwithstanding the mission orbit strategy and the asteroid
97 body shape, accounting for autonomy challenges as the limited computa-
98 tional resources and data storage available on-board. DRL is the chosen
99 method, since it offers both the advantages of a light implementation and
100 generalizing capabilities. The main contribution of this work is the design

101 and development of a DRL-based policy for autonomous decision making of
102 the choosing image acquisition epochs, that increases the efficiency of as-
103 teroid mapping and enhances the shape model reconstruction. The actual
104 benefits for the mapping process are evaluated by comparison between the
105 DRL policy and benchmark policies. The efficiency of the proposed method
106 is evaluated with performance indexes, and its applicability in a real oper-
107 ational scenario with navigation uncertainties is studied. Some preliminary
108 work has been shown in [27], where the method benefits on body shape recon-
109 struction image processing algorithms have been analysed by reconstructing
110 the small body shape with the simulated collected images.

111 The paper is structured as follows. In Section 2 a description of the ex-
112 ploration planning problem in terms of a POMDP is provided, along with
113 the adoption of DRL methods for its solution. In Section 3 the proposed
114 DRL approach for images collection optimization during the body mapping
115 operations is presented. Then, Section 4 deals with the training of the DRL
116 policies. In Section 5 the presented results show a successful policy obtained
117 for images selection, and in Section 6 the policy robustness and computa-
118 tional cost are evaluated, proving to be suitable for an on-board application.
119 Finally, in Section 7 conclusions are drawn.

120 **2. Autonomous Exploration Planning framework**

121 This section deals with the problem of planning under uncertainty, pro-
122 viding the general framework under which small bodies autonomous mapping
123 falls. In the robotics field, autonomous exploration of an unknown environ-
124 ment is typically formulated with an active SLAM approach, coupling the

125 tasks of mapping, localization and planning. Active SLAM can be seen as an
126 instance of POMDPs. The mathematical formulation of POMDPs is briefly
127 introduced and the active SLAM problem is presented as a general model
128 for robotic exploration. Finally, the adopted solution approach with DRL
129 algorithms is detailed.

130 2.1. Partially Observable Markov Decision Processes

131 Markov Decision Processes (MDP) are based on Markov chains, i.e. stochas-
132 tic processes with no memory. This means that the process randomly evolves
133 from one state s_k to another s_{k+1} with a transition probability that depends
134 only on the pair (s_k, s_{k+1}) and not on previous states. The decision-maker
135 (called *agent*) can choose between several possible actions. The transition
136 probability to the next state depends on the chosen action and can be asso-
137 ciated to a scalar *reward*. The agent goal is to maximize the rewards over
138 time, with an optimal *policy*.

139 The absence of memory in MDPs is defined by the *Markov property*:
140 the next state depends only on the current state and action and not on
141 past actions and states, hence the future is conditionally independent of
142 the past, given the present state. This property is essential to many solution
143 algorithms [28]. In real applications, the Markov property requirement can be
144 difficult to meet: state information must be rich enough so that the observed
145 state transition does not depend on additional historical information.

146 When the state is only partially observable, the problem can be defined
147 as a POMDP. In this case the agent can have only a partial knowledge of the
148 environment: the state is not observable but a signal stochastically related
149 to it is. Hence a POMDP can be described as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{T}, \Omega, \mathcal{O} \rangle$,

150 where: \mathcal{S} is the state space, \mathcal{A} is the action space, $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{R}$ is the
 151 immediate reward (often named also cost function), and Ω is the space of
 152 possible observations. A transition probability $\mathcal{T}(s_{k+1} | a_k, s_k)$ governs the
 153 process by mapping a state-action pair to a probability distribution of states
 154 at the next time instant. $\mathcal{O}(o_{k+1} | a_k, s_{k+1})$ is the probability of making
 155 observation o_{k+1} at the next time step, given action a_k that leads to state
 156 s_{k+1} .

157 In the majority of applications POMDPs are computational intractable,
 158 therefore it is better to reduce them to find a computationally tractable solu-
 159 tion. A POMDP can be reduced to a MDP including the agent history h as
 160 internal state. The history is composed by all past actions and observations,
 161 hence history at time step k will be $h_k = \langle a_0, o_1, a_1, \dots, a_{k-1}, o_k \rangle$. The
 162 problem is usually tackled with a less direct approach known as *belief-space*
 163 MDP. This formulation is a tuple $\langle \mathcal{B}, \mathcal{A}, \mathcal{R}_B, \tau \rangle$, where:

- 164 • \mathcal{B} is the belief space, with belief $b_k = p(s_k | h_k)$ equal to the probability
 165 of being in state s after history h .
- 166 • \mathcal{A} is the action space as in the original POMDP.
- 167 • r_B is the expected immediate reward $\mathcal{B} \times \mathcal{A} \rightarrow \mathcal{R}_B$
- 168 • $\tau(b_{k+1} | a_k, b_k)$ is the belief transition function, i.e. the probability of
 169 reaching the new belief b_{k+1} , starting from b_k and performing action
 170 a_k .

171 The optimal policy maximizes the reward in the long term, assuming to act
 172 according to that policy:

$$\pi^* = \operatorname{argmax}_{\pi} \mathbb{E}_{\pi} \left[\sum_{k=0}^{\infty} \gamma^k \mathcal{R}(a_k, b_k) \right] \quad (1)$$

173 This is called also *infinite horizon* problem, since the reward is maximized
 174 over the entire agent lifetime, considering a discount factor $\gamma \in [0, 1]$.

175 2.2. Active Simultaneous Localization And Mapping

176 SLAM consists in estimating a map of an unknown environment and
 177 simultaneously localizing - in the same environment - the moving object: in-
 178 deed, localization is the task of estimating the robot position and orientation
 179 (pose) while moving in the environment [29, 30, 31].

180 SLAM problem can be formulated as follows. The environment map at
 181 time step k is made of a set of n landmarks $\mathbf{l}_k = \{\mathbf{l}^1, \mathbf{l}^2, \dots, \mathbf{l}^n\}$, where
 182 \mathbf{l}^i is the position vector of the i -th landmark. The robot pose \mathbf{x}_k changes
 183 while the robot moves under the control \mathbf{u}_k and can be estimated through
 184 observations of landmarks location \mathbf{z}_k .

185 Active SLAM adds to the SLAM problem the planning task [14, 15, 17].
 186 POMDP provides a framework to investigate the effects of actions and ob-
 187 servations on the agent’s environment perception, thus allowing designing
 188 policies that optimize the agent’s interaction with the environment in some
 189 of its aspects.

190 Since the environment is stochastic, the problem can be described in prob-
 191 abilistic terms according to the belief-space MDP formulation presented in
 192 the previous section 2.1. The state vector is composed by robot pose and
 193 landmark locations $\mathbf{s}_k = (\mathbf{x}_k, \mathbf{l}_k)$ and its belief is $\mathbf{b}_k = p(\mathbf{s}_k | \mathbf{z}_{0:k})$. The ac-
 194 tions that the agent can take coincide with the control $\mathbf{a}_k = \mathbf{u}_k$. Current

195 belief is estimated from past control, past belief, and current observations:
 196 $\mathbf{b}_{\mathbf{k}+1} = \tau(\mathbf{u}_{\mathbf{k}}, \mathbf{b}_{\mathbf{k}}, \mathbf{z}_{\mathbf{k}+1})$. The reward is usually modeled in terms of an ob-
 197 jective function. For instance, a planner can have multiple objectives like
 198 maximizing either coverage or map accuracy while minimizing navigation
 199 duration, motion cost or resources utilization. Several criteria exist to for-
 200 mulate the objective function to be optimized by the planning policy [32, 15].
 201 In particular, the exploration problem consists in choosing the sensing tra-
 202 jectory to obtain the best map.

203 *2.3. Deep Reinforcement Learning*

204 POMDPs are usually computationally intractable. Hence, a reduction
 205 of state, action and policy spaces is needed. An Artificial Neural Network
 206 (ANN) is used to approximate the Q-value, i.e. the expected return over
 207 time. Therefore the optimal policy is the one that maximizes the Q-value:

$$\pi^* = \underset{\pi}{\operatorname{argmax}} Q_{\pi}(\mathbf{a}_{\mathbf{k}}, \mathbf{s}_{\mathbf{k}} | \theta_{\pi}) \quad (2)$$

208 where θ_{π} are the ANN weights and biases. Different approaches can be ap-
 209 plied to formulate the problem with a neural network. This work investigates
 210 two alternatives, NFQ and DQN.

211 The NFQ algorithm scheme is reported in Fig. 1. NFQ sees its major
 212 difficulty in the training data collection: the problem must be suited to be
 213 solved with a random policy that allows the agent-environment interaction
 214 and the collection of the state-action-state triples. If a random policy poorly
 215 performs the environment exploration, the net will be trained only on a subset
 216 of the different situations it could encounter. This approach is model-free,
 217 stable, data efficient and simple to implement.

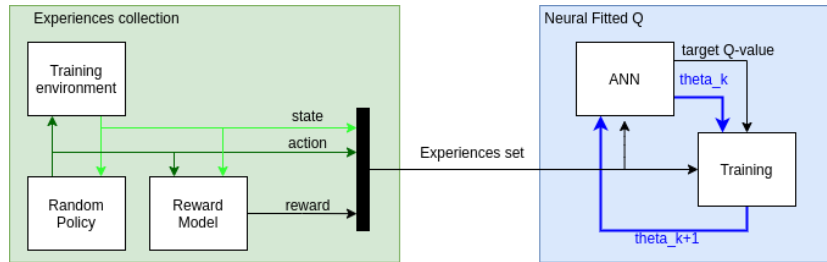


Figure 1: NFQ scheme.

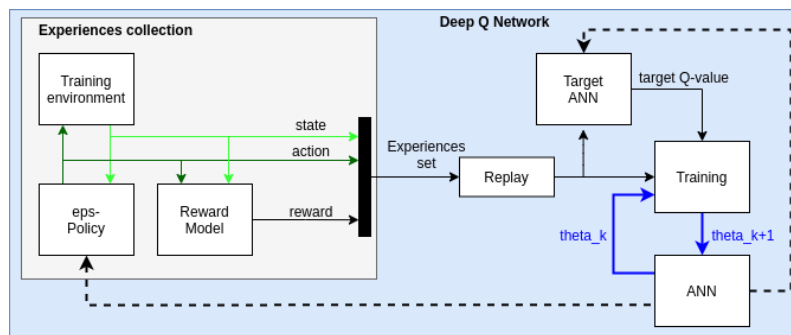


Figure 2: DQN scheme.

218 The DQN algorithm scheme is reported in Fig. 2. Experiences are col-
 219 lected playing many episodes, during which actions are chosen according to
 220 an ϵ -greedy policy. This means that with probability ϵ the action is random
 221 and with probability $(1 - \epsilon)$ the action is the one that maximizes the current
 222 Q-function. The choice of the greedy parameter can be critical to correctly
 223 collect transitions, as also exploration of unknown regions of the state space
 224 is important.

225 3. Images collection planning with Deep Reinforcement Learning

226 The autonomous DRL-based decision making proposed approach is pre-
 227 sented in this section. Then, the detailed definition of the POMDP reduced

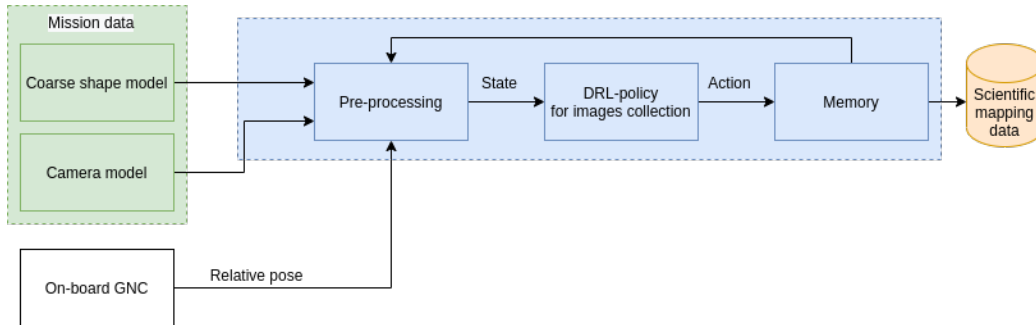


Figure 3: Autonomous DRL-based planning and shape reconstruction validation method.

spaces is provided, in terms of reward \mathcal{R} , belief state space \mathcal{B} and action
space \mathcal{A} .

3.1. Proposed architecture

This paper develops a method to autonomously plan the timing of ob-
servations during the mapping of an unknown small body, with particular
application to imaging for SPC. The goal is to define a policy that improves
mapping quality, while both limiting the amount of images to downlink and
fastening the mapping process. The planning framework is defined as a
POMDP, proposing a novel problem architecture focused on data collection.
DRL is exploited to design the planning policies. A scheme of the proposed
architecture for small bodies imaging and shape reconstruction is schema-
tized in Fig. 3.

Algorithm architecture is designed to be mission-independent and com-
putationally light to cope with limited on board resources. In particular, the
algorithm needs information on the camera Field of View (FoV), the relative
pose between camera and the target, the illumination conditions and the
body low resolution polyhedral shape model, which is already available dur-

245 ing the considered mission phases. Then, data are pre-processed along with
 246 history information of already collected images. The next block is related
 247 with the autonomous decision making: if the current observation epoch is
 248 worth, then the image is taken and actions are recorded. The decision making
 249 problem is solved by means of DRL, dealing with the challenge of optimizing
 250 images collection for small bodies shape reconstruction. The spacecraft acts
 251 according to a policy for next best step selection, i.e. for selecting the most
 252 proper time instant for collecting a new image, based on the relative pose
 253 between camera and body and on the illumination conditions.

254 3.2. Reward definition

255 In this section the reward space \mathcal{R} is defined. SPC benefits from images
 256 with large illumination variation and small viewing angle variation. Scores
 257 related to the photometric angles (see Fig 4) can be defined to assess the
 258 quality of the taken pictures for the shape reconstruction process [25],[27].

259 The overall score S^i associated to the i -th facet of the polyhedral shape
 260 model is given by the weighted sum of five different contributions:

$$S^i = w_1 S_i^i + w_2 S_e^i + w_3 S_{\Delta e}^i + w_4 S_{\Delta \alpha}^i + w_5 S_{\Delta \beta}^i \quad (3)$$

261 where S_i^i is the incidence score, S_e^i the emission one, $S_{\Delta e}^i$ the emission vari-
 262 ation score and $S_{\Delta \alpha}^i$ and $S_{\Delta \beta}^i$ the solar and spacecraft azimuth angle scores.
 263 Such scores are dependent on the spacecraft position and orientation, which
 264 also determines which facets are in view. The mapping resolution is not con-
 265 sidered here, assuming to apply the policy at each mapping stage, thus not
 266 significantly varying the distance from the body. They are defined as follows,
 267 starting from previous studies on the SPC mapping quality [10], [25].

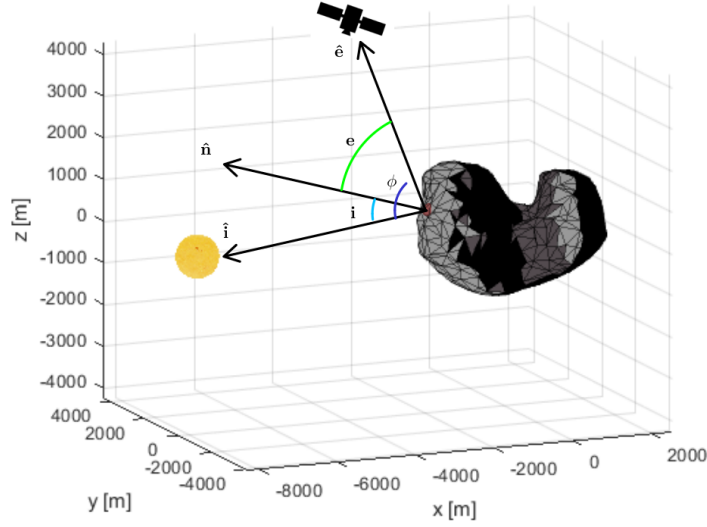


Figure 4: Photometric angles: emission angle e , phase angle ϕ , inclination angle i .

Incidence score. The incidence angle i should be kept between $20^\circ - 60^\circ$ to avoid shadows and excessive brightness, that won't allow the extraction of useful information. Let's define the incidence score S_i^i :

$$s_i = \begin{cases} 1 & \text{if } 20^\circ \leq i \leq 60^\circ \\ \frac{1}{10}i - 1 & \text{if } 10^\circ \leq i \leq 20^\circ \\ -\frac{1}{10}i + 7 & \text{if } 60^\circ \leq i \leq 70^\circ \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$S_i^i = \mu_j(s_i) \quad (5)$$

where μ_j is the mean performed over all the n_{img} taken pictures that contain the facet.

$$\mu_j(x) = \frac{1}{n_{img}} \sum_{j=1}^{n_{img}} (x_j) \quad (6)$$

Emission score. The emission angle should be kept between $10^\circ - 50^\circ$. Hence in a similar manner the emission score S_e^i is defined as follows:

$$s_e = \begin{cases} 1 & \text{if } 10^\circ \leq e \leq 50^\circ \\ \frac{1}{5}e - 1 & \text{if } 5^\circ \leq e \leq 10^\circ \\ -\frac{1}{10}e + 6 & \text{if } 50^\circ \leq e \leq 60^\circ \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$S_e^i = \mu_j(s_e) \quad (8)$$

Emission variation score. Also, a large variation of emission angles is considered beneficial, therefore the emission variation score is:

$$S_{\Delta e}^i = \mu_j \left(\max_k \frac{2\Delta e_{jk}}{\pi} \right) \quad (9)$$

268 where $\Delta e_{jk} = |e_j - e_k|$. Hence for each emission angle e_j under which the
 269 i-th facet is seen, the maximum difference between the considered angle e_j
 270 and all the other angles e_k under which the facets was observed is computed.
 271 Then, all the maximum differences are normalized of $\frac{\pi}{2}$, i.e. the maximum
 272 possible emission variation, and the mean is performed.

Solar and spacecraft azimuth score. Finally, the variation of solar azimuth angles α should be large and the one of spacecraft azimuth angles β small. The respective scores are computed in a similar fashion.

$$S_{\Delta \alpha}^i = \mu_j \left(\max_k \frac{\Delta \alpha_{jk}}{\pi} \right) \quad (10)$$

$$S_{\Delta \beta}^i = 1 - \mu_j \left(\max_k \frac{\Delta \beta_{jk}}{\pi} \right) \quad (11)$$

273 Please note that in this case the normalizing value is π .

274 According to the images history, the facet mapping index m^i is defined
 275 for the i -th facet:

$$m^i = S^i \min \left(1, \frac{n_{img}}{N_{img}} \right) \quad (12)$$

276 where n_{img} and N_{img} are respectively the number of taken images and the
 277 number of ideally necessary images (equal to at last 3 for SPC). The index
 278 m^i can assume values in the interval $[0, 1]$ and in particular the maximum
 279 value represents an ideal perfect mapping.

280 The immediate reward depends on both states and actions:

$$r_k = r_k(s_k, a_k) \quad (13)$$

281 If no action is taken the reward is null. Whenever an image is collected
 282 in a forbidden state $s \in S^-$, a negative reward equal to -1 is returned to
 283 the agent and the image is not accounted for in the successive mapping.
 284 Forbidden states correspond to situations in which either the image is in
 285 complete shadow or the ideal number of images is overcome, occurrence which
 286 might potentially cause problems in on-board data storage. The ultimate
 287 goal is to maximize the mapping index, therefore if the picture is taken in
 288 allowed states the reward is:

$$\tilde{r} = \mu_m \left(\frac{m_k^i - m_{k-1}^i}{m_k^i} \right) \quad (14)$$

289 where m_k^i is the mapping index of facet i at time k and μ_m stands for the
 290 mean over all the facets in the current frame.

Summarizing, the overall reward is:

$$r_k = \begin{cases} -1 & \text{if } a_k = 1 \text{ and } s_k \in S^- \\ 0 & \text{if } a_k = 0 \\ \tilde{r} & \text{otherwise} \end{cases} \quad (15)$$

291 If the agent immediately takes all photos allowed to be sent on ground, along
 292 the successive time steps it will be forced to accept either a zero or a negative
 293 reward. On the other hand, the long term reward will be higher if images
 294 are collected only when it is worth, hence smoothly distributed in time.

295 3.3. Action space

In this section the action space \mathcal{A} is defined. The agent interacts with the environment only by choosing its sensing locations, hence by collecting images, without controlling its relative pose with respect to the body surface. The action at time step k is boolean:

$$a_k = \begin{cases} 0 & \text{if no picture is taken} \\ 1 & \text{otherwise} \end{cases} \quad (16)$$

296 The number η of pictures to be ideally taken in a certain storage time $T_{storage}$
 297 is fixed. After this storage time images are downlinked and therefore the
 298 memory is empty again. The discrete time steps at which an action can
 299 be taken are defined in number equal to the ideal number of images times
 300 a control parameter Δ_c . Ideally with a large control parameter the final
 301 performance would be better, but the number of decisions to be taken would
 302 be too high, entailing a longer and more difficult learning process. Hence a
 303 trade off between performance and learning must be done for the choice of
 304 Δ_c .

305 3.4. State space

306 In this section the belief state space \mathcal{B} is defined. States have been de-
307 signed to synthesize only the information necessary and useful for decision
308 making. In particular, the state is constituted by the *memory state*, *map*
309 *state* and *angles state*. A total of 12 states are defined.

310 *Memory state*. The memory state provides information on the time lapse and
311 number of collected images. The idea is that in a certain time interval T_{storage}
312 pictures can be stored in the on-board memory before being sent to ground.
313 The ideal number of images to communicate at every time interval is η . In
314 particular, the percentage of time spent in the current storage interval and
315 the number of pictures taken n with respect to the ideal number η are fed to
316 the net. The parameters T_{storage} and η can be tuned depending on mission
317 constraints without affecting the algorithm. These inputs help in evaluating
318 how the collection of a new image would impact on data storage.

319 *Map state*. The map state provides general information on the mapping cam-
320 paign advancement. The fraction of area in light of the surface portion in
321 view, which relates to the area percentage whose knowledge will actually be
322 improved by a new picture, can be roughly computed as the ratio between
323 the image facets in light and the total number of facets visible in the image.
324 The map state also includes the mean of the mapping index and its standard
325 deviation over the surface in view and the same quantities computed over
326 the whole body. These data are useful to make a decision on whether the
327 exploration of the area under exam is worth from the coverage point of view.

328 *Angles state.* The angles state gives local information about photometric
329 angles under which the facets in view and in light are seen at the specific
330 epoch. In particular, the angle state includes inclination and emission scores
331 mean, over all the facets in view and in light for the angles of the current time
332 instant only. While the other states are the facets mean of the maximum
333 variation of current Sun azimuth, spacecraft azimuth and emission angles
334 with respect to the angles of already take pictures. These inputs concur in
335 evaluating the possible improvement of SPC for what concerns stereo angles
336 and illumination conditions.

337 The use of statistical quantities (mean and standard deviation) is the only
338 solution that allows keeping constant the number of observed states despite
339 of the change of number of facets in view. Moreover, to understand how the
340 mapping campaign is proceeding, the whole history of past actions should be
341 part of the states as well. Of course to include the whole history in the states
342 observation is unfeasible, but anyway the POMDP is reduced by making
343 part of the history observable. The POMDP is also simplified by assuming
344 the belief equal to the actual state $\mathbf{b}_k \simeq \mathbf{s}_k$. As a future improvement to
345 overcome the drawbacks of classical DQN, the prioritized experience replay
346 could be employed as in [33].

347 *3.5. Neural Network architecture*

348 The architecture of the ANN used to approximate the Q-value function
349 is kept light to achieve a low on-board computational time: a multi-layer
350 perceptron with 2 hidden layers of 10 neurons each is adopted. The network
351 graph is shown in Fig. 5. Such network can be defined a Deep Neural Network
352 [34].

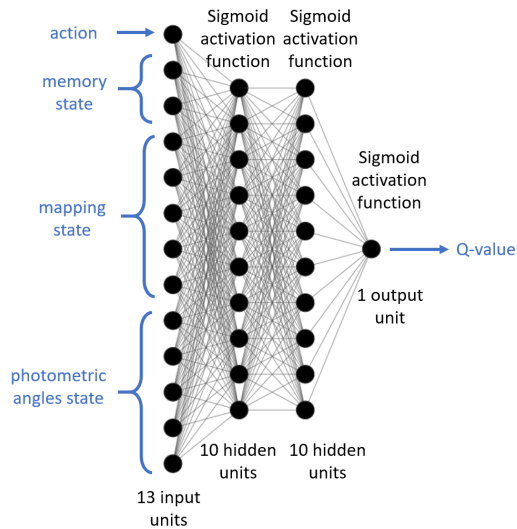


Figure 5: Neural Network architecture.

353 The ANN architecture is the same for both NFQ and DQN and is kept
 354 as simple as possible, with a 13 elements input vector and a scalar output, in
 355 accordance to the necessities of the planning model. Weights and biases of the
 356 ANN are changed during the learning process with resilient back-propagation
 357 (RPROP) steps [35].

358 4. DRL policy training

359 In this section the learning environment is described and the training
 360 results are shown.

361 4.1. Training environment

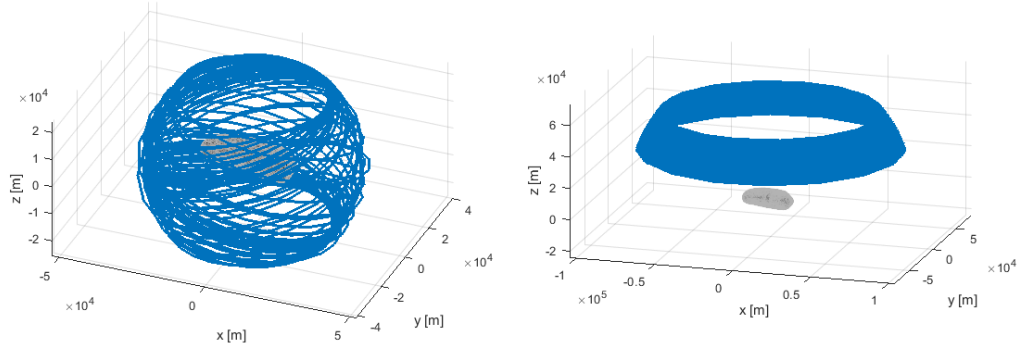
362 To properly define the environment with which the agent interacts during
 363 the learning is fundamental for the learning success. The set of experiences
 364 should be complete, i.e. it should be an exhaustive collection of all possible

365 cases that the agent may encounter. Please note that the state is defined to
366 be independent from the asteroid, orbit and camera characteristics. There-
367 fore, a complete training set is not a set built considering several asteroids
368 and orbits, but a set of examples that sufficiently explores the state space
369 and includes relevant experiences to get to the final goal. Ideally it should
370 contain a whole *mapping stage*, from the beginning to the end, in which all
371 pictures are taken with the same instrument and at about a constant distance
372 from the asteroid. Since the resolution of the maps to be created is finer than
373 that already achieved at larger distances, the stage can be considered inde-
374 pendent from past stages for what concerns the coverage of the asteroid. In
375 a few words, the learning environment should allow the agent to collect both
376 experiences in prohibited states S^- , to identify them and avoid them in the
377 future, and to make very successful actions for mapping. For these reasons,
378 to enhance the learning process, a somewhat unrealistic situation is selected
379 as learning environment:

- 380 • Non-keplerian orbit around asteroid Eros in Figure 6.
- 381 • Camera FoV of 10° .

382 Such scenario allows a great variation of the spacecraft-Sun-body relative
383 geometry.

384 The chosen asteroid is Eros being one of the few shape models publicly
385 available in databases and because its elongated shape allows imaging differ-
386 ent percentages of the body surface, keeping the distance fixed. In fact, the
387 percentage of surface in view varies between 6.4% and 0.7% with a mean of
388 the 2.4% along the orbit.



(a) Spacecraft position - body-fixed frame. (b) Sun direction - body-fixed frame.

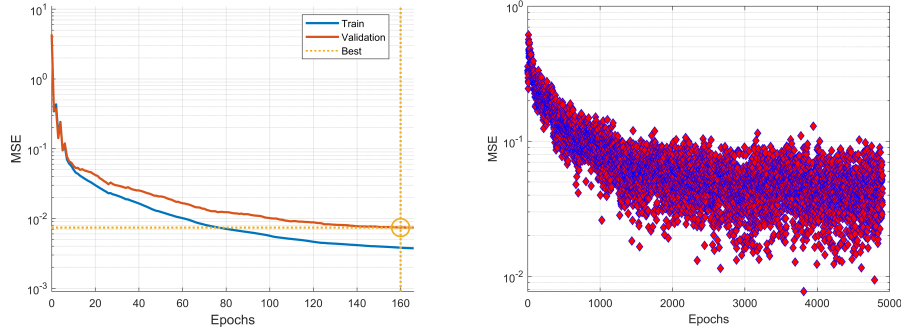
Figure 6: Spacecraft and Sun position in training environment.

389 The trajectory has been obtained considering spherical harmonics pertur-
 390 bations, starting from an initial condition corresponding to osculating orbital
 391 parameters of null eccentricity, 45° inclination and radius twice the asteroid
 392 maximum one. For what concerns the body illumination, some areas remains
 393 always in shadow, as it can be seen from Sun direction in the body-fixed frame
 394 shown in Figure 6. The reason is that the body spin axis inclination with
 395 respect to the ecliptic north is larger than Eros orbit inclination. This allows
 396 to frequently collect also negative experiences for the body mapping, which
 397 need to be learned and avoided.

398 The training simulation environment also assumes that the ideal number
 399 of images during one episode is 500, with an ideal frequency of 1 picture per
 400 hour. The data downlink and control parameters are:

401 • $T_{\text{storage}} = 10 \text{ h}$

402 • $\eta = 10$



(a) Training Mean Square Error for NFQ network. (b) Training Mean Square Error for DQN network.

Figure 7: DRL policies learning process.

403 • $\Delta_c = 3$

404 In practice, the orbit is discretized so that the number of points in the storage
 405 time is three times the number of photos allowed. Hence the control interval
 406 between one action and the next is quite coarse. This interval can be refined
 407 in future works.

408 4.2. Learning process

409 RPROP is selected as training algorithm because of its robustness. In
 410 particular, batch learning is preferred to incremental learning, because the
 411 training set has a low dimension (500 ideal images and $\Delta_c = 3$ lead to a total
 412 number of 1500 experiences). Input and output scaling is performed on the
 413 whole experiences set.

414 *NFQ learning.* The Mean Square Error between network outputs and targets
 415 is reported for one training iteration in Figure 7a, where one epoch corre-
 416 sponds to a weights update step on the entire set of experiences. As it can

417 be observed, the Mean Square Error smoothly decreases until the validation
418 check is met, i.e. the validation error stops decreasing. As expected, the
419 training error is lower than the validation error.

420 *DQN learning.* The Mean Square Error evolution during the learning is re-
421 ported in Figures 7b. In this case one epoch corresponds to one RPROP step
422 on the mini-batch. As it can be noticed, Mean Square Error decreases but it
423 is much less stable with respect to the NFQ case. This is a consequence of
424 the different batch sizes used in the two algorithms.

425 **5. DRL policy performance**

426 In this section, the performance of the DRL policy, trained with DQN
427 and NFQ methods, is presented. The two DRL methods are compared to
428 benchmark policies in different mission scenarios to verify their generalizing
429 capability, which is of great importance when exploring an unknown envi-
430 ronment.

431 *5.1. Benchmarks and performance metrics definition*

432 *Benchmarks.* A first numerical validation is performed by comparing the
433 DRL-based algorithm with two different simple benchmarks: a policy that
434 takes pictures at regular intervals (UNI) and another that randomly selects
435 the image acquisition instants (RAND). UNI takes a picture every Δ_c time
436 steps. For the RAND strategy if $n_k > n_{k,UNI}$ the image is discarded and all
437 presented results are the mean over 100 runs. For UNI, NFQ, and DQN only
438 1 run is needed, since they are deterministic policies.

439 *Performance indexes.* Often DRL results are compared just with the nu-
 440 merical final score obtained during the episode. In such a way however, to
 441 critically analyze how policies actually behave is hard. Being the design and
 442 learning procedure highly based on engineering judgment, test results are
 443 presented not with final reward scores, but by means of some complemen-
 444 tary indexes that facilitate the performance understanding:

- 445 1. Final number of collected images I_n , (the lower the better).
- 446 2. Final mapping index $I_{map} = \mu_M(m_{k_{end}}^i)$, where μ_M is the mean over all
 447 the body facets (the higher value the better).
- 448 3. Integral mapping index over the campaign $I_{sum} = \frac{1}{\Delta_c} \sum_k \mu_M(m_k^i)$, (the
 449 higher the better).

450 Such parameters quickly allow verifying whether the modeled reward actually
 451 leads to an improvement of the proposed tasks: data reduction and mapping
 452 enhancement and fastening.

453 5.2. Test cases definition

454 Test cases have been chosen to cover all the relevant aspects for the algo-
 455 rithm application. In particular, four different bodies are considered: Eros,
 456 on which the training has been performed, Itokawa, that presents an elon-
 457 gated shape, Bennu, with diamond shape, and 67P-CG, with two-lobes shape.
 458 Small bodies are assumed on Keplerian orbits around the Sun, with constant
 459 spin axis orientation and rotational period. The camera is assumed to have
 460 a conical FoV of 3° and a shape model of 1000 facets is considered available
 461 on-board. The length of each test episode is set to 1500 time steps, with
 462 $\Delta_c = 3$, and an ideal final images number of $1500/\Delta_c = 500$.

463 Sensitivity analysis run for the above mentioned small bodies by varying
464 the following quantities:

- 465 • The distance from the body, that affects both relative dynamics and
466 percentage of surface in the camera FoV. In particular, the quantity
467 here referred to is the *interest ratio*, i.e. the ratio between distance
468 from the body center and maximum body radius.
- 469 • The body rotational period T , that influences illumination conditions
470 variation and again relative pose.
- 471 • The orbit inclination i , that changes the surface portion object of the
472 mapping.

473 5.3. Detailed results for 67P test case

474 For the sake of brevity, results are presented in this section for the 67P
475 scenario only, among the ones above mentioned. In fact, it well represents
476 challenges linked to an extremely irregular body shape, as self-shadowing
477 and self-occlusion, and it can be assumed as a worst-case example for the
478 achieved performances. Please note that the presented case differs from the
479 case exploited for the learning process.

480 The basic simulation scenario for comet 67P has the following parameters:

- 481 • Rotational period $T_{rot} = 12.4$ h.
- 482 • Circular polar orbit, with interest ratio equal to 6.
- 483 • Percentage of surface in view in the range 0.3% to 2.9%.

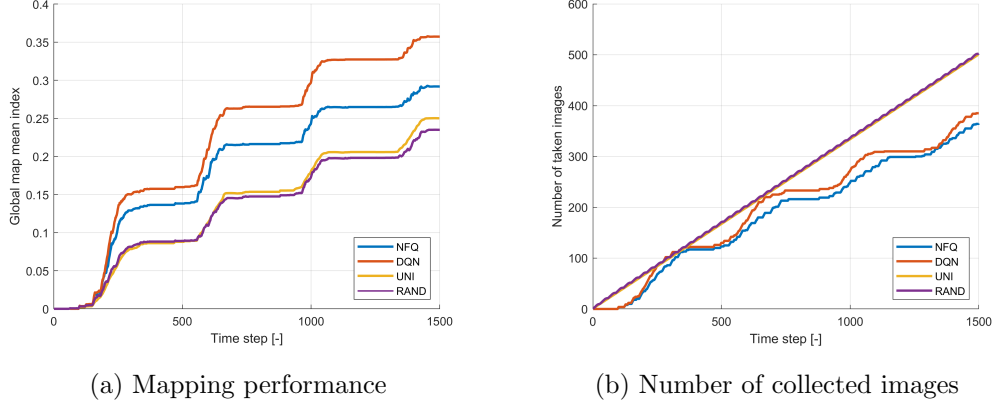


Figure 8: Polar orbit at 67P. Comparison of the different strategies.

484 The evolution of the mapping quality index and the number of collected
 485 images are respectively shown in Fig. 8 for a circular polar orbit at 67P, with
 486 interest ratio 6. For both NFQ and DQN, not only the amount of collected
 487 data is equal or less with respect to RAND and UNI strategy, but also the
 488 mapping index is higher.

489 The final mapping index is shown for NFQ, DQN and UNI strategies in
 490 Fig. 9. In this case the mapping is hindered by Sun illumination but also by
 491 the significant self-shadowing and self-occlusion.

492 5.4. Sensitivity analysis results

493 Detailed results of the sensitivity analysis are here reported for the Eros
 494 and 67P scenarios, respectively in Table 1 and Table 2. For most test cases
 495 in the four bodies scenarios DQN proves to be the best policy. In some cases,
 496 I_{map} has a similar value for the three strategies, but the goal is achieved faster
 497 by NFQ and DQN, that have a larger I_{sum} . A general trend observed is that
 498 with NFQ the number of collected pictures is lower than DQN and UNI, but

Table 1: Eros, sensitivity analysis.

	i = 30 deg			i = 60 deg			i = 90 deg		
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.21	66.87	500	0.22	69.10	500	0.22	69.43
NFQ	430	0.25	87.64	431	0.26	87.58	404	0.26	86.50
DQN	498	0.24	83.30	493	0.26	84.91	434	0.29	89.67
	Interest Ratio = 6			Interest Ratio = 8			Interest Ratio = 10		
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.33	114.38	500	0.38	142.34	500	0.41	160.01
NFQ	381	0.33	120.64	377	0.37	143.31	488	0.39	150.86
DQN	317	0.37	135.90	320	0.38	151.13	326	0.39	159.87
	T = 2 h			T = 5 h			T = 12 h		
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.21	57.78	500	0.22	61.80	500	0.20	58.97
NFQ	409	0.25	79.00	402	0.25	84.18	415	0.25	74.06
DQN	459	0.28	81.41	446	0.29	86.20	469	0.25	71.76

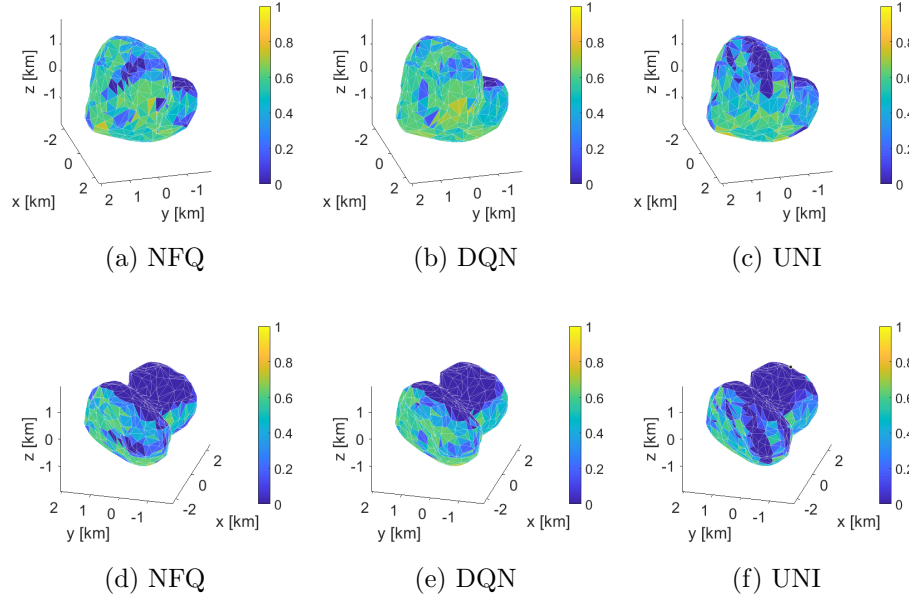


Figure 9: 67P, facets final mapping index.

499 with DQN a larger mapping performance is achieved, sometimes even with
 500 a lower number of pictures.

501 As visible in Table 1 and Table 2, by increasing the interest ratio the
 502 two DRL strategies become less efficient: even if I_n is reduced, I_{map} and
 503 I_{sum} are comparable to the UNI strategy. This trend has been observed for
 504 all considered small bodies and may be due to two reasons: the percentage
 505 of surface in view is out of training interval and of typical mission values;
 506 or when a large portion of the body is imaged it is more difficult to have
 507 control on the viewing conditions of all facets in the frame. In fact 1% of the
 508 surface means to consider about 10 facets, while 10% corresponds to 100:
 509 very different viewing conditions may be present in the same picture. Please
 510 note that in any case the number of pictures for DQN and NFQ is largely

Table 2: 67P-CG, sensitivity analysis.

	i = 30 deg			i = 60 deg			i = 90 deg		
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.27	79.67	500	0.27	72.15	500	0.25	67.01
NFQ	356	0.28	95.70	342	0.30	95.24	365	0.29	91.55
DQN	476	0.30	94.65	417	0.36	112.69	385	0.36	111.34
Interest Ratio = 8 Interest Ratio = 10 Interest Ratio = 12									
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.33	102.68	500	0.41	125.66	500	0.40	130.36
NFQ	310	0.32	114.30	290	0.40	126.50	279	0.40	126.52
DQN	285	0.38	134.45	285	0.41	143.01	314	0.41	133.63
	T = 2 h			T = 5 h			T = 12 h		
	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}	I_n	I_{map}	I_{sum}
UNI	500	0.21	50.94	500	0.25	62.84	500	0.24	65.59
NFQ	355	0.31	87.28	354	0.29	93.19	379	0.26	84.03
DQN	398	0.36	98.93	381	0.36	110.49	427	0.28	95.00

511 reduced in spite of a small difference in I_{map} and I_{sum} .

512 DRL-policy has been trained in a completely different scenario, concern-
513 ing both body shape and orbit, and has been designed to be easily employed
514 into a wide variety of different mission scenarios. Therefore, an optimal
515 behaviour is actually not expected to be reached, even is a significant en-
516 hancement in scientific mapping and data collected, compared to simpler
517 acquisition strategies, is sought.

518 Sensitivity analyses results show that the proposed solutions are capable
519 to deal with far-off different scenarios and outperform the UNI and RAND
520 benchmarks. The presented algorithm can work independently of the relative

521 dynamics between spacecraft and small body, proving to be highly flexible: in
522 all considered small bodies scenarios the policy for selection of the observation
523 times actually enhances the efficiency of data collection.

524 **6. Compatibility with on-board application**

525 The applicability of the proposed architecture for an on-board application
526 is analysed in this Section. In particular, two relevant aspects are studied: the
527 robustness of the DRL policy to uncertain inputs coming from the on-board
528 navigation system, and the computational effort required by the architecture.

529 *6.1. DRL policy robustness to uncertainty*

530 During close proximity operations, the on-board knowledge of the relative
531 pose with respect to the body surface is limited by the navigation accuracy.
532 This aspect directly influences the inputs to the DRL-policy and may lead
533 to a behaviour different from the expected, since uncertain inputs were not
534 considered at all during the training. For these reasons, some tests are per-
535 formed on the DRL-policy, introducing errors in the knowledge of the relative
536 pose of the spacecraft.

537 The considered testing scenario consists in a quasi-Keplerian circular or-
538 bit at asteroid Bennu, at 2.5 km distance and with an inclination of 45°. The
539 assumed camera has a FoV of 10°. In particular, 6 tests are performed con-
540 sidering an increasing uncertainty separately for position and pointing, which
541 are perturbed by additive Gaussian white noise with standard deviation σ .
542 Finally, test 7 considers both effects, with the uncertainty expected for the
543 study case, i.e. 100 m for the relative position and 1° for the pointing. To

544 have a fair term of comparison, in each simulation the spacecraft ground-
545 truth pointing and position history is the same and only the state belief
546 is different. Each test has been run with 100 simulations; mean value and
547 standard deviation of the mapping index and acquired frames are reported
548 in Table 3. The mapping index is reported as a percentage with respect
549 to the complete mapping, which is impossible to achieve, and the acquired
550 frames are reported as percentage of the imposed maximum capability of
551 data storage and communication.

Table 3: DRL-policy performance with uncertainties on relative state.

Test	Position	Pointing	Mapping	Frames
ID	σ [m]	σ [°]	index [%] (σ)	acquired [-] (σ)
ALL	0	0	48.3 (-)	150 (-)
0	0	0	45.9 (-)	66 (-)
1	100	0	43.9 (1.4)	68 (4)
2	250	0	37.1 (4.7)	67 (5)
3	500	0	20.9 (5.7)	70 (3)
4	0	1	45.8 (0.5)	71 (5)
5	0	3	47.4 (0.3)	82 (7)
6	0	5	47.9 (0.4)	92 (8)
7	100	1	44.1 (0.1)	70 (4)

552 Table 3 shows that the DRL-policy is quite robust to both kinds of uncer-
553 tainties and still achieves a good performance within the typical navigation

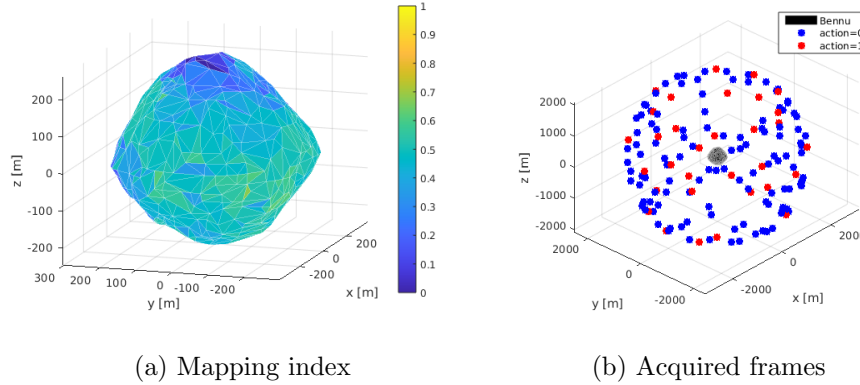


Figure 10: Benu, DQN performance without uncertainties (test 0).

554 uncertainties (test #7), but the performance decreases when the state belief
 555 is very far from the real state.

556 Test ALL is performed to assess the maximum ideally achievable mapping
 557 without any constraint on the data storage (UNI policy applied with high
 558 sampling frequency), while test #0 assesses the behaviour of the DRL policy
 559 with a perfect knowledge of the state (see Fig. 10).

560 Tests #1-3 show that the number of frames acquired is comparable to
 561 the test-0, but the actual mapping performance decreases with a higher po-
 562 sition uncertainty. Since the DRL-policy assumes to point Benu centre of
 563 mass from the belief of its relative position, the information regarding to
 564 the illumination of the surface in view is affected, leading to a lower quality
 565 mapping.

566 Tests #4-6 highlight a different behaviour of the net: a larger number
 567 of images is collected, increasing the mapping quality thanks to the larger
 568 amount of data. In test #6 almost all the images are collected: the belief
 569 of the current mapping is worse than the actual one because part of the

570 body is believed to exit the FoV; thus the policy continues collecting data to
571 complete the body coverage even if they are not necessary.

572 In test #7 the policy confirms to be robust to combined uncertainties in
573 the state estimation, leading to a good mapping of the object, but with a
574 rise of data acquired with respect to absence of uncertainties. The considered
575 uncertainties are in line with a realistic scenario and the AI-policy is verified
576 to outperform a classical UNI scheduling in terms of amount of data and in
577 relation to images quality.

578 *6.2. Computational cost preliminary assessment*

579 A computational analysis is addressed to determine the feasibility and
580 limits of a possible on-board implementation of the algorithm. The algorithm
581 is implemented in MATLAB and all tests are run on an Intel[®] Core[™] i7-
582 5500U CPU, clocked at 2.4 GHz, paired to a 16 GB DDR3 memory.

583 *Computational time for a global mapping case.* The computational time to
584 take a single decision is evaluated over 500 runs - i.e. the ideal number of
585 images for an episode - considering a 1000 facets spherical shape model and
586 a typical case in which 1.5% of the surface is in view. Results in Table 4 and
587 Fig. 11 show that a low computational time is required, with a mean value
588 of 33.5 ms.

589 *Surface in view: percentage variation.* Another analysis was performed, to
590 examine the computational cost trend with respect to the percentage of sur-
591 face in view. All other parameters are kept as before. The mean time linearly
592 increases with the surface portion, as shown in Fig. 12, where each point is
593 the mean computational time over 500 decisions. The mean time can be

Table 4: Computation time to take a single decision, with 1.5% of the surface in view, 500 runs.

Time [ms]	
Average	33.5
Minimum	7.4
Maximum	132.5

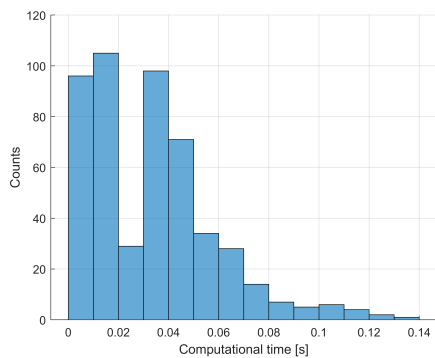


Figure 11: Computation time to take a single decision. Time with 1.5% of the surface in view, 500 runs.

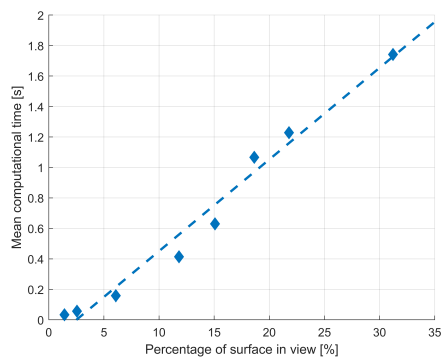


Figure 12: Computation time to take a single decision, varying the surface portion in view.

594 about two orders of magnitude larger when the surface portion in view is
595 above the 30 %.

596 Implementation on flight hardware would corresponds to increased com-
597 putation time, potentially introducing a certain delay in the decision making.
598 The effect of such a delay would be the imaging of a slightly different area
599 with respect to the expected one, due to the target rotation and the spacecraft
600 dynamics. For a hypothetical fast rotating spherical body with 2 h rotational
601 period, the surface displacement in 1 s is of about the 0.1 % of the character-
602 istic dimension. If the surface portion in view is large, such displacement is
603 likely not significant; while for a small area in view the computational time
604 is minimum. Similar considerations hold considering the specific spacecraft
605 dynamics.

606 **7. Conclusion**

607 In conclusion, an AI-based planning policy for enhancing the mapping of
608 an asteroid or comet is here proposed. Achieved results of the presented ap-
609 proach reveal the methodology to be a promising step forward in autonomous
610 operations, helping in decreasing the human effort during the mapping phases
611 of unknown small bodies and increasing imaging exploitation efficiency with
612 a simple and flexible scheme. The merits of the proposed architecture are
613 the decoupling of the decision-making process from spacecrafts dynamics, the
614 autonomy improvement with very low risks for the mission, and the general
615 validity of the planning framework, which is mission-independent and does
616 not require learning during operations. The DRL-based strategies generaliz-
617 ing capabilities are verified through numerical simulations, obtaining promis-

618 ing results. Two results are consequent to the application of the DRL-based
619 policy: an increased performance mapping efficiency and a correct handling
620 of memory storage during the mapping. The strategy robustness to uncer-
621 tain inputs coming from the on-board navigation is tested, confirming its
622 suitability for a realistic scenario. Future work will further investigate the
623 effectiveness of the proposed techniques with more challenging benchmarks
624 cases.

625 **References**

- 626 [1] R. W. Gaskell, O. S. Barnouin-Jha, D. J. Scheeres, A. S. Konopliv,
627 T. Mukai, S. Abe, J. Saito, M. Ishiguro, T. Kubota, T. Hashimoto,
628 J. Kawaguchi, M. Yoshikawa, K. Shirakawa, T. Kominato, N. Hirata,
629 H. Demura, Characterizing and navigating small bodies with imaging
630 data, *Meteoritics and Planetary Science* 43 (2008) 1049–1061.
- 631 [2] B. Giese, J. Oberst, R. Kirk, W. Zeitler, The topography of asteroid ida:
632 A comparison between photogrammetric and two-dimensional photocli-
633 nometric image analysis, *International Archives of Photogrammetry and*
634 *Remote Sensing* 31 (1996) B3.
- 635 [3] A. Capannolo, F. Ferrari, M. Lavagna, Families of bounded orbits near
636 binary asteroid 65803 didymos, *Journal of Guidance, Control, and Dy-*
637 *namics* 42 (2019) 189–198.
- 638 [4] S. B. Broschart, G. Lantoine, D. J. Grebow, Quasi-terminator orbits
639 near primitive bodies, *Celestial Mechanics and Dynamical Astronomy*
640 120 (2014) 195–215.

- 641 [5] D. J. Scheeres, Orbit mechanics about asteroids and comets, *Journal of*
642 *Guidance, Control, and Dynamics* 35 (2012) 987–997.
- 643 [6] D. Scheeres, B. Sutter, A. Rosengren, Design, dynamics and stability
644 of the osiris-rex sun-terminator orbits, *Advances in the Astronautical*
645 *Sciences* 148 (2013) 3263–3282.
- 646 [7] C. Circi, A. D’Ambrosio, H. Lei, E. Ortore, Global mapping of asteroids
647 by frozen orbits: The case of 216 kleopatra, *Acta Astronautica* 161
648 (2019) 101–107.
- 649 [8] Y. Zhang, X. Zeng, F. Zhang, Spacecraft hovering flight in a binary
650 asteroid system by using fuzzy logic control, *IEEE Transactions on*
651 *Aerospace and Electronic Systems* 55 (2019) 3246–3258.
- 652 [9] X. Zeng, S. Gong, J. Li, K. T. Alfriend, Solar sail body-fixed hovering
653 over elongated asteroids, *Journal of Guidance, Control, and Dynamics*
654 39 (2016) 1223–1231.
- 655 [10] T. A. Pavlak, S. B. Broschart, G. Lantoine, Quantifying mapping orbit
656 performance in the vicinity of primitive bodies (2015).
- 657 [11] S. B. Broschart, D. J. Scheeres, Control of hovering spacecraft near small
658 bodies: application to asteroid 25143 itokawa, *Journal of Guidance,*
659 *Control, and Dynamics* 28 (2005) 343–354.
- 660 [12] T. Wen, X. Zeng, C. Circi, Y. Gao, Hop reachable domain on irregularly
661 shaped asteroids, *Journal of Guidance, Control, and Dynamics* 43 (2020)
662 1269–1283.

- 663 [13] R. P. de Santayana, M. Lauer, Optical measurements for rosetta navi-
664 gation near the comet, in: Proceedings of the 25th International Sym-
665 posium on Space Flight Dynamics (ISSFD), Munich.
- 666 [14] H. J. S. Feder, J. J. Leonard, C. M. Smith, Adaptive mobile robot nav-
667 igation and mapping, *The International Journal of Robotics Research*
668 18 (1999) 650–668.
- 669 [15] H. Carrillo, I. Reid, J. A. Castellanos, On the comparison of uncertainty
670 criteria for active slam, in: 2012 IEEE International Conference on
671 Robotics and Automation, IEEE, pp. 2080–2087.
- 672 [16] T. Kollar, N. Roy, Trajectory optimization using reinforcement learning
673 for map exploration, *The International Journal of Robotics Research* 27
674 (2008) 175–196.
- 675 [17] A.-a. Agha-mohammadi, S. Chakravorty, N. M. Amato, FIRM :
676 Sampling-based Feedback Motion Planning Under Motion Uncertainty
677 and Imperfect Measurements, *The International Journal of Robotics*
678 *Research* 33 (2003) 268–304.
- 679 [18] D. Izzo, M. Märten, B. Pan, A survey on artificial intelligence trends
680 in spacecraft guidance dynamics and control, *Astrodynamics* 3 (2019)
681 287–299.
- 682 [19] B. Gaudet, R. Furfaro, R. Linares, Reinforcement learning for angle-
683 only intercept guidance of maneuvering targets, *Aerospace Science and*
684 *Technology* 99 (2020) 105746.

- 685 [20] Y. Zhou, E.-J. van Kampen, Q. Chu, Incremental model based online
686 heuristic dynamic programming for nonlinear adaptive tracking control
687 with partial observability, *Aerospace Science and Technology* 105 (2020)
688 106013.
- 689 [21] M. Riedmiller, Neural fitted q iteration – first experiences with a data
690 efficient neural reinforcement learning method, in: J. Gama, R. Ca-
691 macho, P. B. Brazdil, A. M. Jorge, L. Torgo (Eds.), *Machine Learning:*
692 *ECML 2005*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp.
693 317–328.
- 694 [22] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G.
695 Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski,
696 et al., Human-level control through deep reinforcement learning, *Nature*
697 518 (2015) 529.
- 698 [23] S. Li, P. Cui, H. Cui, Autonomous navigation and guidance for landing
699 on asteroids, *Aerospace science and technology* 10 (2006) 239–247.
- 700 [24] H. Shang, X. Wu, D. Qiao, X. Huang, Parameter estimation for op-
701 timal asteroid transfer trajectories using supervised machine learning,
702 *Aerospace Science and Technology* 79 (2018) 570–579.
- 703 [25] V. Pesce, A.-a. Agha-mohammadi, M. Lavagna, Autonomous navigation
704 & mapping of small bodies, in: *2018 IEEE Aerospace Conference*, IEEE,
705 pp. 1–10.
- 706 [26] D. M. Chan, A.-a. Agha-mohammadi, Autonomous imaging and map-

- 707 ping of small bodies using deep reinforcement learning, in: 2019 IEEE
708 Aerospace Conference.
- 709 [27] M. Piccinin, M. R. Lavagna, Deep reinforcement learning approach for
710 small bodies shape reconstruction enhancement, in: AIAA Scitech 2020
711 Forum, p. 1909.
- 712 [28] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction,
713 MIT press, 2018.
- 714 [29] H. Durrant-Whyte, T. Bailey, Simultaneous localization and mapping:
715 part i, IEEE Robotics Automation Magazine 13 (2006) 99–110.
- 716 [30] T. Bailey, H. Durrant-Whyte, Simultaneous localization and mapping
717 (slam): part ii, IEEE Robotics Automation Magazine 13 (2006) 108–
718 117.
- 719 [31] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira,
720 I. Reid, J. J. Leonard, Past, present, and future of simultaneous local-
721 ization and mapping: Toward the robust-perception age, IEEE Trans-
722 actions on Robotics 32 (2016) 1309–1332.
- 723 [32] L. Mihaylova, T. Lefebvre, H. Bruyninckx, K. Gadeyne, J. De Schut-
724 ter, A comparison of decision making criteria and optimization methods
725 for active robotic sensing, in: International Conference on Numerical
726 Methods and Applications, Springer, pp. 316–324.
- 727 [33] J. Jiang, X. Zeng, D. Guzzetti, Y. You, Path planning for asteroid hop-
728 ping rovers with pre-trained deep reinforcement learning architectures,
729 Acta Astronautica 171 (2020) 265–279.

- 730 [34] A. Géron, Hands-on machine learning with Scikit-Learn and Tensor-
731 Flow: concepts, tools, and techniques to build intelligent systems,
732 O'Reilly Media, Inc., 2017.
- 733 [35] M. Riedmiller, H. Braun, A direct adaptive method for faster backpropa-
734 gation learning: The rprop algorithm, in: IEEE international conference
735 on neural networks, IEEE, pp. 586–591.