AURORAS: Automated Remote Observation for Recognition and Analysis of Safe Landing Sites via Deep Learning

Luísa M. C. Santos
LIAREA, SolarAI Group and PGCAP
INPE
São José dos Campos, Brazil
luisa.santos@inpe.br

Matheus C. Domingos
LIAREA and PGCAP
INPE
São José dos Campos, Brazil
matheus.domingos@inpe.br

Hércules C. S. Pereira
LIAREA and PGCAP
INPE
São José dos Campos, Brazil
hercules.pereira@inpe.br

Rafael D. C. Santos
PGCAP
INPE
São José dos Campos, Brazil
rafael.santos@inpe.br

Luis Eduardo A. Vieira
SolarAI Group
INPE
São José dos Campos, Brazil
luis.vieira@inpe.br

Valdivino A. Santiago Júnior
LIAREA and PGCAP
INPE
São José dos Campos, Brazil
valdivino.santiago@inpe.br

Elcio H. Shiguemori
PGCAP
IEAv
São José dos Campos, Brazil
elcio@ieav.cta.br

***Abstract*—The autonomous identification of safe landing sites on small celestial bodies with highly irregular terrains is a challenge for deep space missions. Traditional methods often rely on active sensors (e.g., LiDAR) and 3D reconstructions based on Digital Elevation Models (DEMs), which impose mass, power, and complexity constraints that are impractical for many small-body missions. To address this limitation, this study proposes AURORAS (Autonomous Remote Observation for Recognition and Analysis of Safe Sites), a supervised semantic segmentation framework designed to detect safe and unsafe regions directly from high-resolution 2D RGB images (2048×2048 pixels, 2–5 cm/pixel) acquired by the Rosetta OSIRIS Narrow Angle Camera. A dataset of 17 manually labeled images was prepared and divided into non-overlapping 256×256 patches, initially totaling 1,088 samples. To improve model generalization, on-the-fly data augmentation was applied during training, including random horizontal and vertical flips, and rotations of 90°, 180°, and 270°. Three convolutional neural network (CNN) architectures were evaluated: U-NetResNet18, SegNet, and UNetResNet18 + SegNet with an Attention Mechanism. All models were trained in two stages, pretraining on augmented patches and fine-tuning on the original data, using a systematic split into training, validation, and test sets. Evaluation metrics included Pixel Accuracy, Intersection over Union (IoU), Precision, Recall, and F1-Score. Among the tested architectures, SegNet achieved the best overall performance, making it particularly suitable for minimizing false positives in critical space applications. The proposed method demonstrates the feasibility of using passive optical sensing and deep learning for autonomous landing site selection without relying on DEMs or active sensors.**

***Index Terms*—Automated Remote Observation, Deep Learning, Semantic Segmentation, Comet, Landing, Sites Selection.**

LIAREA - Laboratório de Inteligência ARtificial para Aplicações AeroEspaciais e Ambientais; PGCAP - Programa de Pós-Graduação em Computacão Aplicada; INPE - Instituto Nacional de Pesquisas Espaciais; IEAv - Instituto de Estudos Avançados.

I. Introduction

Autonomous landing on small celestial bodies, such as asteroids and comets, remains a complex challenge in interplanetary exploration. These environments present extreme topographical irregularities, unpredictable gravitational fields, and limited prior terrain information, which significantly hinder the ability of onboard systems to perform real-time, safe landing operations. Traditional hazard detection and avoidance (HDA) systems typically rely on active sensors such as LiDAR or radar to construct Digital Elevation Models (DEMs) that support slope and obstacle analysis [1], [3]. While these approaches offer accurate 3D reconstruction capabilities, they increase the mission's mass, power requirements, and system complexity factors that are often impractical for deep space or small-body missions [4], [5].

In response to these constraints, recent studies have explored the use of passive optical systems combined with deep learning to directly identify safe landing regions from 2D images [7]. These image-based approaches benefit from reduced resource consumption and high-resolution spatial data but introduce new computational challenges, such as the need to extract hazard-relevant features, like boulders and slope, without explicit depth information [2]. Diaz et al. [7] and Tomita et al. [1], for instance, demonstrates the promise of semantic segmentation models in detecting hazardous regions on lunar surfaces from monocular imagery.

Semantic segmentation has emerged as a powerful tool in this domain due to its capacity to classify every pixel in an image according to predefined terrain safety criteria. Architectures like U-Net [8], SegNet [9], and attention-based variants such as Attention U-Net [10] have been successfully adapted from biomedical imaging to space applications.

Recent benchmarking studies [2] confirm that these models achieve competitive accuracy for planetary hazard detection and are suitable for onboard deployment with optimization.

Additionally, deep learning techniques are increasingly being integrated into relative navigation pipelines for autonomous spacecraft guidance [11], where feature extraction from visual data becomes essential in the absence of GNSS. This convergence of navigation and perception underscores the relevance of developing robust, lightweight hazard detection systems based solely on camera imagery.

Despite their potential, deep learning models for hazard detection in space missions face several practical limitations. As highlighted in recent surveys [11], these methods rely heavily on large labeled datasets, which are scarce in the context of planetary exploration. The extreme environmental conditions, such as variable illumination, absence of atmospheric scattering, and strong shadows—further complicate the task, impacting the generalization ability of convolutional models. Moreover, onboard computational constraints demand lightweight and efficient architectures capable of real-time inference. In this scenario, hybrid approaches that combine deep learning with classical geometric methods, such as Extended Kalman Filters (EKF) and perspective-based pose estimation, have shown greater robustness and interpretability in mission-critical systems.

Simulated studies on lunar terrains using high-fidelity Digital Terrain Models (DTMs), such as those conducted with U-Net in Blender-rendered environments, have demonstrated high segmentation accuracy even under six degrees of freedom (6-DOF) in camera movement [7]. These models show promising results under favorable solar illumination conditions, yet their performance degrades when images are overexposed due to high solar incidence, highlighting the dependency of image-based approaches on lighting conditions. Such results emphasize the trade-offs involved in deploying semantic segmentation models in real missions, where generalization under variable lighting remains a critical challenge.

In parallel, classical AHDA strategies remain highly relevant. The Chinese Chang'e-3 mission, for example, implemented a two-stage hazard detection scheme that combined coarse 2D image analysis at high altitudes with precise 3D terrain reconstruction via LiDAR at lower altitudes [12]. Candidate landing sites were then evaluated using spiral search algorithms and safety metrics that considered both topographic flatness and control energy for maneuver execution. While current proposals suggest more intelligent onboard systems including flash LiDAR and infrared sensors for all-weather operations, these remain computationally intensive and are still being matured for real-time embedded deployment.

In this context, we propose AURORAS (Autonomous Remote Observation for Recognition and Analysis of Safe Sites), a supervised semantic segmentation framework designed to identify safe and unsafe regions directly from high-resolution 2D optical images (2048x2048 pixels with 2–5 cm/pixel) acquired by the OSIRIS Narrow Angle Camera onboard the Rosetta mission. Rather than introducing a new algorithm, AURORAS presents a methodological contribution that leverages CNN architectures, U-NetResNet18, SegNet, and a hybrid UNetResNet18 + SegNet with an Attention Mechanism, to demonstrate the feasibility of passive, image-based terrain evaluation for landing site selection.

The methodology involves patch-wise segmentation of 17 manually labeled OSIRIS images, extensive on-the-fly data augmentation, and two-stage training to enhance generalization. This study contributes to the field by reinforcing the viability of vision-only autonomous landing systems in environments where active sensors are not deployable, thus aligning with recent trends toward AI-driven autonomy in deep space missions [6].

II. Literature Review

A. Asteroid and Comet Missions: Emerging Interest

In recent years, global interest in missions targeting comets and asteroids has increased significantly, driven by both scientific curiosity and strategic potential. This growing attention is reflected in the increasing number of initiatives led by space agencies, academic institutions, and private companies dedicated to exploring the technical, economic, and legal feasibility of asteroid missions. Recent developments in missions such as Lucy, Psyche 16, Hayabusa-2, and OSIRIS-REx demonstrate a clear progression toward higher technological readiness for asteroid mining and space resource utilization [20].

B. Hazard Detection for Planetary Landing

Traditionally, hazard detection systems for planetary landings have relied on active sensors like Light Detection and Ranging (LiDAR) or radar, optical remote sensing technology that measures properties of reflected light in order to measure distance, combined with the processing of DEMs to evaluate slope, roughness, and obstacle presence. However, active sensors often involve significant mass, power, and complexity constraints, making them unsuitable for many small-body missions.

Recent advances have focused on vision-based hazard detection using passive optical sensors and deep learning techniques that can infer landing safety directly from visual patterns without heavy reliance on explicit 3D models [1], [2].

C. CNN-Based Optical Navigation and Real-Time Inference

The integration of convolutional neural networks (CNNs) within visual navigation systems has become a pivotal aspect of autonomous spacecraft operations, thereby enhancing their efficacy. Convolutional neural networks (CNNs) have been shown to be particularly useful for tasks such as hazard mapping, landing site detection, and pose estimation directly from camera imagery [21]. By leveraging onboard optical data, CNNs obviate the necessity for bulky active sensors or external tracking infrastructure.

It is noteworthy that CNN-based navigation has been demonstrated to offer real-time viability, attributable to its expeditious inference times and minimal memory requirements. These attributes are of paramount importance in the context

of space platforms that are constrained in their resources. Conventional systems frequently employ geometric modelling (e.g., Kalman filters), however, CNNs present an alternative approach that operates from input to output, learning direct mappings from imagery to position or terrain safety classes. It is evident from a survey of the relevant literature that, despite the possibility of hybridising CNNs with classical estimation models, there is a clear preference for standalone CNN pipelines in the context of optical navigation [21], [22].

D. Semantic Segmentation for Hazard Detection

Deep learning-based semantic segmentation has recently emerged as a powerful tool for terrain understanding and hazard detection. Landing missions on small bodies, such as asteroids and comets, remain extremely challenging, with only a handful successfully conducted so far — for instance, just five missions to asteroids have been carried out to date. In this context of limited references and operational difficulties, it is important to highlight the complexity involved not only in landing site selection but also in spacecraft descent and control [4]. Recent studies, such as [1] and [2], have demonstrated the strong performance of CNNs like U-Net and DeepLabV3 in classifying safe and unsafe landing zones, even in environments characterized by noise and high uncertainty.

E. Visual Navigation in Irregular Terrains

Beyond segmentation, visual odometry and terrain-relative navigation techniques have been explored. Notably, Johnson et al. [3] developed structure-from-motion algorithms for autonomous hazard detection using monocular cameras, and Mancini et al. [4] proposed siamese CNNs for asteroid navigation.

F. Rosetta Mission

The Rosetta mission, conducted by the European Space Agency (ESA), was designed to study comet 67P/Churyumov–Gerasimenko. The ten-year mission aimed primarily to deploy a lander to the surface in order to study the composition and characteristics of the soil. The Rosetta mission's landing of Philae on comet 67P underlines the importance of autonomous hazard detection. Despite manual site selection, Philae experienced multiple bounces due to unexpected surface irregularities [5]. The challenges encountered during the Philae landing, arising from the underestimation of surface slope and obstacle density, underscore the imperative for more rigorous, automated terrain analysis frameworks, such as the one proposed in this study.

G. Landing Site Selection Process in the Rosetta Mission

One of the major challenges was selecting a safe site for the Philae lander. The selection process was highly complex, developed between August and November 2014, and involved image analysis, flight dynamics trajectories, terrain modeling, and risk assessment. The process followed the steps outlined below [5]:

1) **Launch of Rosetta**
   The Rosetta spacecraft was launched to orbit the comet and collect data to support the landing planning.
2) **OSIRIS - Comet Imaging**
   The OSIRIS instrument (Optical, Spectroscopic, and Infrared Remote Imaging System) was used to capture high-resolution images of the comet. These images enabled the identification of topographic features such as large boulders and craters.
3) **DTM - Digital Terrain Modeling**
   Based on the images, a Digital Terrain Model (DTM) was created. This model allowed for the analysis of the comet's surface relief, the identification of flat and rugged areas, and the calculation of local slopes.
4) **Geodetic Reference Frame Emulation on the Comet**
   Since the comet has an irregular shape, it was necessary to establish a local coordinate system to consistently reference areas of interest. This system specifically built for the comet functioned as a "GPS" based on the EME2000 reference frame. This reference is used for deep space missions, the commonly used Earth-centered inertial (ECI) frame is defined with the Earth's Mean Equator (EME) at 12:00 Terrestrial Time on 1 January 2000.
5) **Analysis for the Philae Lander**
   Mapping of potential landing sites for the Philae lander was carried out. The criteria considered included the availability of sunlight for more than 6.2 h for recharge the batteries by solar panel - analysis until 6 months later.
6) **Measurement of Terrain Slope Magnitude**
   The candidate areas needed to have slopes less than 30° (z axis of the lander relative to normal terrain of the comet) to prevent Philae from tipping over upon landing. Slope maps were generated from the DTM to eliminate dangerous regions.
7) **Risk Assessment Related to Rocks and Slope**
   Boulders larger than 1m in diameter were considered hazardous. Risk assessment was conducted using Monte Carlo simulations - 10^4 trajectories, where multiple simulated landing points were generated to estimate the probability of collision with obstacles. So was putted one "buffer" of security with more 1m of diameter around the rocks.

Despite these extensive preparations, Philae's landing highlighted the difficulty of hazard detection in small-body environments. A mechanical failure during touchdown — specifically the malfunction of anchoring harpoons and thruster — caused the lander to bounce and settle in an unintended, poorly illuminated location approximately one kilometer from the target [5].

Using high-resolution 2D images offers several advantages for deep space missions. They enable lightweight and passive sensing, minimizing payload mass and energy consumption. Their fine spatial resolution allows for the detection of critical

surface features such as rocks and craters and eliminate the necessity for complex 3D reconstructions during hazard assessment [7].

III. Methodology

The AURORAS methodology is shown in Fig. 1. In the sequence, we describe its main steps.

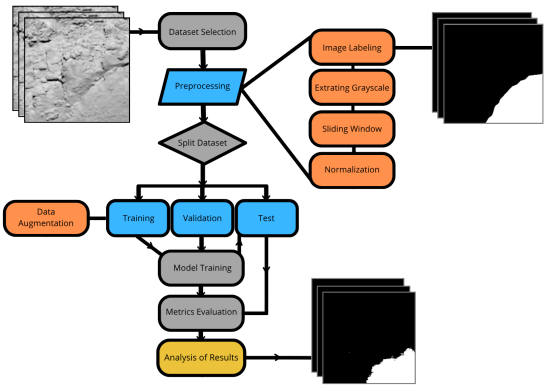

Fig. 1. The AURORAS methodology.

A. Dataset Creation

The selection of the dataset was conducted using the Planetary Science Archive (PSA) of the European Space Agency (ESA). From this repository, 17 high-resolution RGB frames with a resolution of 2048×2048 pixels were retrieved from the Rosetta Mission [15]. The selection criteria were applied according to the platform interface illustrated in Fig. 2.

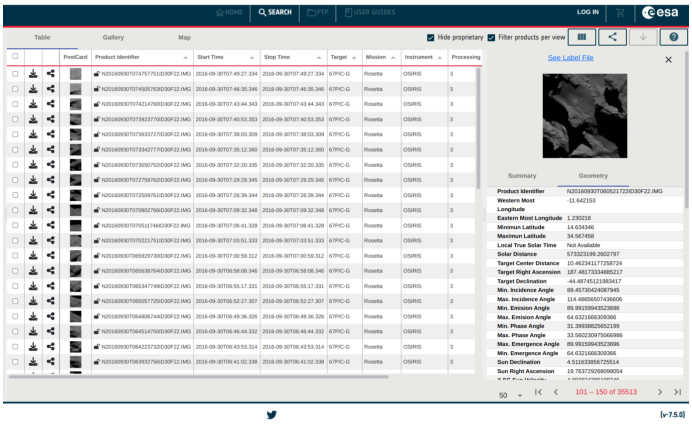

Fig. 2. Plataform of the PSA from ESA [15].

The dataset for analysis was obtained by applying a set of objective filters, exclusively selecting data from the Rosetta mission's orbiter component, specifically at altitudes around 8 to 10 km, targeting comet 67P/Churyumov-Gerasimenko (67P/C-G). Observations utilized were captured by the OSIRIS NAC, strictly involving calibrated Level 3 image data. Further selection criteria restricted the analysis to imagery within the visible (VIS) and near-infrared (NIR) wavelengths, and files identified by the .IMG product ID format, ensuring consistency and scientific rigor in the dataset. Fig. 3 illustrates an example of a selected image frame used in the dataset.

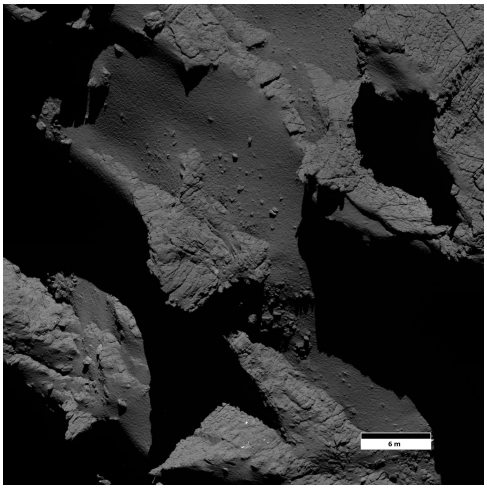

Fig. 3. Comet frame from the dataset [15].

B. Preprocessing and Training Setup

The dataset comprises 17 high-resolution RGB images (2048×2048 px) captured by the OSIRIS Narrow Angle Camera aboard ESA's Rosetta orbiter, with an approximate ground sampling distance of 2–5 cm/px. Each image was manually annotated via Roboflow platform [17], resulting in binary masks indicating landing safety (1 = safe, 0 = unsafe).

The dataset and annotation masks used in this study are publicly available in the AURORAS Dataset repository on Kaggle [18], which contains 17 high-resolution OSIRIS NAC frames from comet 67P, manually labeled for binary segmentation of safe and unsafe areas. The complete training pipeline, including preprocessing, model architecture, and evaluation scripts, is openly accessible as a Kaggle Notebook [19], supporting reproducibility and future extensions of this research.

To construct the patch-based dataset, each image was divided into non-overlapping 256×256 patches, yielding 64 patches per image and a total of 1088 annotated patches. These were split into 50% for training (543), 20% for validation (218), and 30% for testing (327), using stratified sampling to preserve the class distribution, as illustrated in Fig. 1.

Each RGB patch was converted to a single-channel grayscale image to match the network input and emphasize intensity-based terrain features. Pixel values were normalized to the $[0, 1]$ range by dividing by 255. For inference, a sliding window of 256 × 256 with 50% overlap was applied across the test images to generate dense segmentation maps.

Data augmentation was applied exclusively to the training set and executed dynamically during training through a custom PyTorch `Dataset` subclass. The augmentation strategy included random horizontal and vertical flips and rotations of 90°, 180°, and 270°, implemented via the Albumentations library [14]. This on-the-fly approach increases spatial variability across epochs without increasing dataset size or storage

requirements. It also mimics variations in camera orientation during remote observations of irregular celestial bodies [7].

All models were trained for 100 epochs using the AdamW optimizer with a learning rate of 1×10^{-4} and a weight decay of 0.01. A batch size of 8 was used for all runs, balancing memory usage and convergence. The Binary Cross-Entropy with Logits (`BCEWithLogitsLoss`) was chosen as the loss function, given the binary segmentation objective.

To ensure reproducibility, the random seed was fixed at 42 across Python (`random`), NumPy, and PyTorch (`torch.manual_seed(42)` and `torch.cuda.manual_seed(42)`). The implementation was conducted in PyTorch [13].

Training and validation losses were recorded at each epoch, along with accuracy, precision, recall, F1-score, and intersection-over-union (IoU), to monitor learning progress and detect overfitting. Final evaluation was conducted on the held-out test set using the same metrics.

C. Model Architecture and Training

We adopted 3 models for semantic segmentation; the first was a U-Net-style segmentation network with a ResNet-18 encoder pre–trained on ImageNet. After replacing the first convolution to accept a single-channel input, we attach four up–sampling blocks mirroring the encoder's feature maps, concatenating skip–connections at each resolution and finishing with a 1×1 convolution to produce a single–channel logit mask.

The UNetResNetEncoder (Fig. 4) leverages a ResNet-18 [16] backbone, originally composed of an initial 7×7 convolutional layer followed by four stages of two residual blocks each, to form a powerful encoder for grayscale segmentation. To accommodate single-channel inputs, we replace the first 7×7 convolution with a one-channel variant while preserving the original residual blocks (layer1–layer4). Each block employs identity "skip connections" that add its input to the output of two convolutional layers (followed by ReLU), thereby ensuring unhindered gradient flow, mitigating vanishing-gradient issues, and yielding rich, multi-scale feature maps (64 → 512 channels) as spatial resolution is progressively downsampled. In the decoder, these learned features are upsampled via transposed convolutions and fused with their corresponding encoder activations through UNet-style skip connections, which preserves fine-grained spatial details. Each upsampling stage is further refined by paired convolutional blocks (ConvBlock), and the final feature map is restored to the full input resolution via bilinear interpolation followed by a 1×1 convolution, producing a single-channel probability map indicating the presence of the target class at each pixel [8].

The second was the SegNet represented bellow by the Figure 5 that employs a symmetric encoder–decoder design that captures hierarchical features via successive 3×3 convolution, BatchNorm(BN), and ReLU blocks followed by 2×2 max-pooling (which records the pooling indices) to halve spatial resolution while doubling feature channels from 64 up to 512.

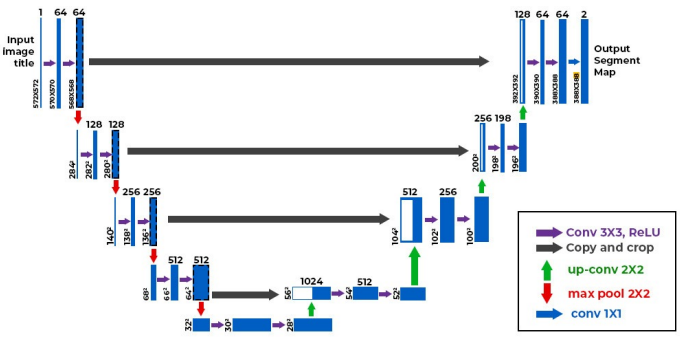

Fig. 4. Unet Architecture [8].

In decoding, the stored indices drive 2×2 MaxUnpool2d layers to invert pooling—reconstructing each feature map to its pre-pooled size by repositioning activations and zero-filling the rest—after which paired 3×3 conv–BatchNorm–ReLU modules halve the channel count back to 64. This index-guided unpooling preserves sharp boundaries and fine details without introducing the extra parameters of learned deconvolutions. A final 1×1 convolution then reduces the feature map to a single-channel logit mask for per-pixel classification [9].

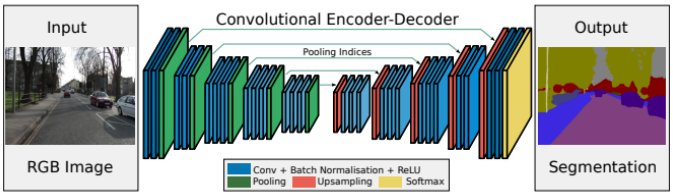

Fig. 5. SegNet Architecture [9].

The third model augments the first two architectures with an AttentionBlock that learns a spatially varying gate to refine encoder features before they feed into the decoder [10]. Specifically, it projects both the coarse decoder signal and the corresponding encoder activation into a low-dimensional space (via 1×1 conv+BN), sums them, applies ReLU, and uses another 1×1 conv+BN+sigmoid to produce a single-channel attention mask that highlights the most relevant regions. This SegNet–U-Net–Attention network then combines SegNet's precise, index-based unpooling, U-Net's skip connections, and attention-gated feature fusion: four downsampling stages (64→128→256→512 channels) each apply two 3×3 conv–BN–ReLU blocks and max-pooling (with indices saved), while four upsampling stages use MaxUnpool2d, two conv–BN–ReLU refinements, and attention-guided concatenation with the matching encoder output, halving channels at each step. A final 1×1 convolution restores the original resolution with a single logit map, ensuring both boundary detail preservation and focused processing on the most informative spatial regions.

D. Metrics

After training completes, the model is evaluated on the held-out test set. We compute global pixel-wise accuracy, precision, recall, F1-score, and intersection-over-union (IoU) by thresholding the sigmoid-activated logits at 0.5.

a) Pixel Accuracy: This metric the proportion of correct predictions relative to the total number of evaluated samples. It is calculated as the sum of true positives (*TP*) and true negatives (*TN*) divided by the total number of samples. This metric provides an overall view of the model's performance but may be affected by class imbalance issues [6].

$$\text{Pixel Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

b) Precision: This metric quantifies the proportion of correctly predicted positive instances relative to the total number of positive predictions, emphasizing the model's reliability in accurately identifying landing strips. Complementarily, recall, also referred to as sensitivity, was used to evaluate the model's ability to correctly identify positive instances. Recall is calculated as the ratio of true positives (*TP*) to the sum of true positives and false negatives (*FN*). This metric is particularly relevant in scenarios where false negatives have significant consequences, such as in environmental monitoring applications [6].

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

c) Recall: Also referred to as sensitivity, this metric was employed to measure the model's ability to correctly identify positive instances, focusing on minimizing false negatives. It is calculated as the number of true positives (*TP*) divided by the sum of true positives (*TP*) and false negatives (*FN*). This metric evaluates the model's sensitivity and is particularly important in scenarios where missing positive cases is costly, such as identifying landing strips in remote sensing applications [6].

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

d) F1-Score: This metric is defined as the harmonic mean of precision and recall, was adopted to provide a balanced assessment of these metrics in scenarios with imbalanced classes. The combined use of these metrics enabled a robust evaluation of the model, taking into account both its overall effectiveness and its ability to handle class imbalance [6].

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

E. Intersection over Union (IoU)

Intersection over Union (IoU), also known as the Jaccard Index, is a widely used evaluation metric in semantic segmentation, particularly suitable for quantifying the spatial overlap between predicted and reference regions for a specific class. In the context of binary classification for landing site detection, IoU measures how well the model predicts one of the two classes—typically the "safe" region by comparing the intersection and union of predicted and actual areas.

Mathematically, the IoU is defined as:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{5}$$

where:

- TP (True Positives): pixels correctly predicted as belonging to the class of interest,
- FP (False Positives): pixels incorrectly predicted as belonging to the class,
- FN (False Negatives): pixels that belong to the class in the ground truth but were not predicted as such.

IoU values range from 0 to 1, where higher scores indicate better overlap between the predicted and reference masks. Unlike overall accuracy, IoU penalizes both false positives and false negatives, making it more reliable in datasets with class imbalance, like terrain segmentation where unsafe regions may dominate the image. In binary configurations, IoU is often reported separately for each class or focused on the "positive" class, depending on the task.

IV. Experiments and Results

This section presents a comprehensive analysis of the semantic segmentation performance achieved by three distinct neural network architectures. Each model's strengths and trade-offs are discussed to identify the optimal architecture for effectively segmenting landing sites, highlighting the impact of different decoding strategies and attention mechanisms on boundary precision and detection sensitivity.

TABLE I
Comparison of the segmentation model results

Model	**PixelAcc**	**Precision**	**Recall**	**F1-score**	**IoU**
UNetResNet Encoder	0.84	0.81	0.87	0.84	0.72
SegNet	**0.88**	**0.82**	0.89	**0.85**	**0.74**
UNet+SegNet +Attention	0.87	0.80	**0.90**	0.84	0.73

Table I shows that all three architectures deliver strong performance, but each excels on slightly different metrics. SegNet achieves the highest overall accuracy (88%) and precision (82%), as well as the best F1-score (85%) and IoU (74%), indicating that its pooling-index–driven decoder effectively preserves boundary details while minimizing false positives. The UNetResNetEncoder follows closely, matching SegNet's recall (87%) and yielding a competitive F1-score (84%) and IoU (72%), thanks to its deep residual backbone and skip connections. The combined UNet+SegNet+Attention model achieves the highest recall (90%), reflecting its superior sensitivity to true positives through the attention gates, though this comes with a slight drop in precision (80%) relative to the other two. Overall, SegNet offers the best balance of precision

and overlap (IoU), while the attention-enhanced variant is most effective at capturing all positive instances, and the ResNet-based UNet strikes a middle ground.

A. UNetResNetEncoder

The model's predictions exhibit signs of under-segmentation in certain regions, especially in the finer structures of unsafe zones. This suggests that although the architecture is capable of learning meaningful features, its performance could be improved with additional training epochs or a larger annotated dataset to better handle complex terrain details. Nevertheless, the UNetResNet18 [8], [16] provides valuable results by effectively identifying major non-landable areas, positioning it as a strong baseline for scenarios with limited data. In safety-critical applications such as autonomous landing site selection, the model offers a promising trade-off between computational efficiency and terrain awareness, serving as a reliable foundation for early-stage segmentation in interplanetary exploration.

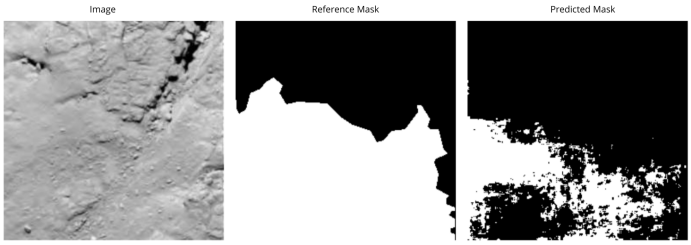

Fig. 6. Result from the Masks of the UNet+ResNet18. From left to right: (i) an image patch; (ii) the reference mask (ground truth); and (iii) the predicted output of the model.

B. SegNet

SegNet [9] proved highly effective at segmenting regions of regular terrain, accurately distinguishing safe zones, free of irregularities or rocks, from unsafe areas. As shown in Fig. 7, the predicted masks faithfully mirror the shape and spatial location of the reference masks, correctly positioning the safe regions over the original image.

However, a slight reduction in the total extent of the predicted safe regions compared to the reference masks is observed. This conservative bias is advantageous for deep-space missions: by prioritizing a smaller set of zones with high certainty of safety, it minimizes the risk of false positives and potential damage to the lander. Thus, SegNet [9] adopts an approach that aligns with the stringent reliability and safety standards required for landing on celestial bodies.

C. UnetResNet18+SegNet+Attention

Figure 8 reveals that the hybrid UNet–SegNet model augmented with an attention mechanism [10] produces noticeably noisier segmentations, introducing scattered artifacts and irregular boundaries that undermine mask coherence. Although it captures the overall contours separating safe and unsafe landing zones, this noise detracts from the morphological integrity and visual clarity of the predictions.

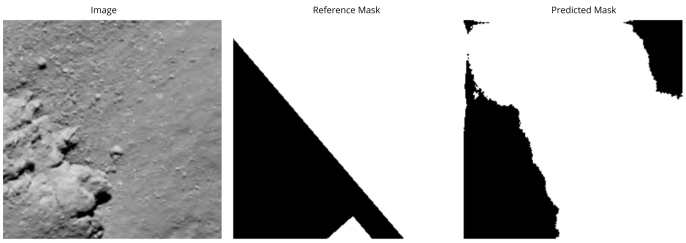

Fig. 7. Result from the Masks of the SegNet. From left to right: (i) an image patch; (ii) the reference mask (ground truth); and (iii) the predicted output of the model.

Moreover, the model shows a propensity for false negatives in genuinely safe regions, suggesting that the attention module overemphasizes local details at the expense of global spatial context. In mission-critical planetary landing scenarios, such a bias can lead to overly conservative exclusion of viable sites. While the hybrid architecture demonstrates enhanced feature learning, further refinement is needed to strike a better balance between sensitivity and precision in safety-critical applications.

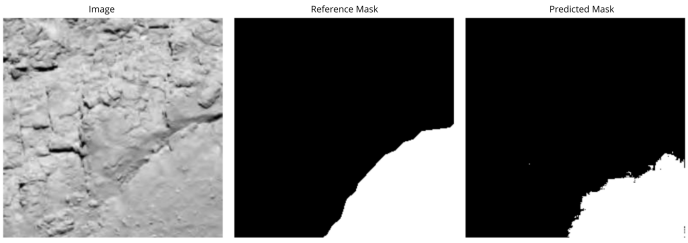

Fig. 8. Result from the Masks of the UnetResNet18+SegNet+Attention. From left to right: (i) an image patch; (ii) the reference mask (ground truth); and (iii) the predicted output of the model.

It is imperative to emphasise that each model was run a single time using a fixed random seed (42) to ensure reproducibility. Therefore, the reported metrics reflect only this single run and do not account for variability from random initialization or data shuffling. While this limits statistical confidence, the results serve as an initial benchmark. Future studies will include multiple runs and statistical analysis to enhance the model's generalization capabilities.

V. Conclusions

The present study investigated the application of deep learning-based semantic segmentation models to identify safe landing sites on the surface of comet 67P/Churyumov–Gerasimenko using high-resolution 2D images from the Rosetta mission. In order to demonstrate the potential and limitations of three distinct architectures (UNetResNet18, SegNet, and a hybrid UNet+SegNet+Attention Mechanism models) under the constraints of limited data and operational safety in space environments, the present study trained and evaluated the architectures.

Among the models evaluated, SegNet achieved the best overall performance, making it particularly suitable for scenarios where minimising false positives is crucial to increasing the chances of mission success. The UNetResNet18 model, while demonstrating slightly reduced accuracy, exhibited notable overlap capabilities with the reference mask, thus serving as a foundation for future exploration and enhancement in conjunction with alternative, more suitable encoder types. The hybrid model, with its heightened focus, exhibited augmented sensitivity; however, it concomitantly engendered augmented segmentation noise and false negatives, a phenomenon with the potential to impede mission adaptability.

In the context of deep space exploration, where mission success frequently depends on accurate terrain assessment and cautious decision-making, these findings underscore the value of conservative and accurate segmentation strategies. It is preferable to identify a smaller number of landing zones with a high degree of confidence than to overestimate the safety of regions and risk mission failure.

In the subsequent phases of this project, we intend to enhance the evaluation protocol. This includes calculating confusion matrices to analyze class-specific errors and detect segmentation biases. Furthermore, the intention is to expand the dataset by incorporating a more extensive range of annotated images and developing a new multi-class dataset. This will assist in mitigating overfitting and enhancing model generalization. Concurrently, Explainable Artificial Intelligence (XAI) techniques will be integrated into the pipeline, for example on Grad-CAM, with the objective of generating heatmaps that highlight the activation areas corresponding to the image regions most relevant to the model's predictions. This methodological approach will provide interpretable visual explanations that can guide the identification of the areas of greatest importance to the model, thus increasing the reliability of the results obtained.

ACKNOWLEDGMENT

This work was supported by the infrastructure provided by the National Institute for Space Research (INPE), funded by the National Council for Scientific and Technological Development (CNPq), grant number [130228/2024-9] and Coordination for the Improvement of Higher Education Personnel (CAPES). This study has been conducted within the SolarAI Group and the *Laboratório de Inteligência ARtificial para Aplicações AeroEspaciais e Ambientais* (LIAREA), both from INPE.

REFERENCES

[1] K. Tomita, T. Nakamura, T. Shirai, and A. Murao, "Hazard Detection Algorithm for Planetary Landing Using Semantic Segmentation," in *IEEE Aerospace Conference*, 2020. doi: 10.1109/AERO47225.2020.9172727.

[2] T. Claudet, A. Dupuy, M. Desvignes, and E. Lemaire, "Benchmark Analysis of Semantic Segmentation Algorithms for Safe Planetary Landing Site Selection," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 58, no. 5, pp. 4452–4465, Oct. 2022. doi: 10.1109/TAES.2022.3169042.

[3] A. E. Johnson, J. F. Montgomery, and L. H. Matthies, "Vision-Guided Landing of an Autonomous Helicopter in Hazardous Terrain," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2005, pp. 3966–3971. doi: 10.1109/ROBOT.2005.1570713.

[4] P. Mancini, G. Frigerio, S. Silvestri, and R. Pieri, "Deep Learning for Asteroids Autonomous Terrain Relative Navigation," in *Acta Astronautica*, vol. 206, pp. 68–79, 2023. doi: 10.1016/j.actaastro.2023.01.024.

[5] E. Jurado, J. L. Cano, and M. R. Sanchez, "Rosetta Lander Philae: Flight Dynamics Analyses for Landing Site Selection and Post-landing Operations," in *Acta Astronautica*, vol. 127, pp. 404–417, 2016. doi: 10.1016/j.actaastro.2016.06.038.

[6] D. M. W. Powers, "Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[7] S. Diaz, J. Jordan, C. Carmichael, A. Schwartz, and E. B. Fitts, "Image-Based Lunar Landing Hazard Detection via Deep Learning," in *Proceedings of the IEEE Aerospace Conference*, 2021, pp. 1–11. doi: 10.1109/AERO50100.2021.9438402.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[9] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[10] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[11] J. Song, D. Rondao, and N. Aouf, "Deep learning-based spacecraft relative navigation methods: A survey," *Acta Astronautica*, vol. 212, pp. 176–197, 2024. doi: 10.1016/j.actaastro.2023.11.013.

[12] Q. Zou, Z. Liu, and Y. Liu, "Innovative hazard detection and avoidance strategy for autonomous safe planetary landing," *Acta Astronautica*, vol. 111, pp. 286–298, 2015. doi: 10.1016/j.actaastro.2015.02.021.

[13] A. Paszke *et al.*, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 8024–8035, 2019.

[14] A. Buslaev, V. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020.

[15] European Space Agency, "Planetary Science Archive (PSA)," *https://archives.esac.esa.int/psa*, accessed Jun. 2025.

[16] A. V. S. Abhishek, V. R. Gurrala, and L. Sahoo, "ResNet18 Model With Sequential Layer for Computing Accuracy on Image Classification Dataset," International Journal of Creative Research Thoughts (IJCRT), vol. 10, no. 5, pp. c176–c181, May 2022. ISSN: 2320-2882.

[17] Roboflow, "Roboflow: Annotate, preprocess, and manage computer vision datasets," 2021. [Online]. Available: https://roboflow.com

[18] L. M. C. Santos, "AURORAS Dataset: Safe and Unsafe Landing Site Labels for Comet 67P," Kaggle, 2025. [Online]. Available: https://www.kaggle.com/datasets/lusasantos/auroras-dataset/data

[19] L. M. C. Santos, "Comet Landing Sites – Deep Learning Semantic Segmentation," Kaggle Notebook, 2025. [Online]. Available: https://www.kaggle.com/code/lusasantos/cometlandingsites

[20] S. Srivastava, S. S. Pradhan, B. Luitel, P. Manghaipathy, and M. Romero, "Analysis of technology, economic, and legislation readiness levels of asteroid mining industry: A base for the future space resource utilization missions," New Space, vol. 11, no. 1, pp. 1–13, 2023. doi: 10.1089/space.2022.0031.

[21] T. Campbell, R. Furfaro, R. Linares, and D. Gaylor, "A Deep Learning Approach for Optical Autonomous Planetary Relative Terrain Navigation," in AIAA Guidance, Navigation, and Control Conference, 2017, doi:10.2514/6.2017-1964.

[22] E. Turan, S. Speretta, and E. Gill, "Autonomous Navigation for Deep Space Small Satellites: Scientific and Technological Advances," Acta Astronautica, vol. 193, pp. 56–74, 2022, doi:10.1016/j.actaastro.2021.12.036.