Deep Learning-Optimized Monocular Navigation for Autonomous Rendezvous and Proximity Maneuvers in Small Satellite Missions

Deep Learning-Optimized Monocular Navigation for Autonomous Rendezvous and Proximity Maneuvers in Small Satellite Missions

Lucrezia Lovaglio
Department of Electronics and Telecommunications
Politecnico di Torino
Turin, Italy
lucrezia.lovaglio@polito.it

Fabrizio Stesina
Department of Mechanical and Aerospace Engineering
Politecnico di Torino
Turin, Italy
fabrizio.stesina@polito.it

***Abstract*—Accurate estimation of the position and orientation of a spacecraft during proximity operations—such as rendezvous, docking, on-orbit servicing (OOS), and active debris removal (ADR)—is critical to ensuring mission success and safety. Traditional visual navigation methods based on hand-engineered feature matching often struggle with robustness and generalization, while existing deep learning approaches face limitations due to heuristic hyperparameter tuning and limited training data. In this work, a novel convolutional neural network (CNN)-based architecture for monocular pose estimation of non-cooperative spacecraft is proposed, specifically designed to improve robustness across diverse operational scenarios. The model is trained on a high-fidelity synthetic dataset comprising approximately 25,000 images, simulating realistic proximity conditions with variations in lighting, background textures, and spacecraft geometries. To assess its performance, an extensive benchmarking study is conducted against representative State-of-the-Art methods using standardized evaluation metrics and controlled test conditions. The results demonstrate the competitive performance of the proposed method and provide critical insights into the factors affecting pose estimation accuracy in realistic spaceborne applications.**

***Index Terms*—Visual Navigation, On-Orbit Servicing, Convolutional Neural Networks, spacecraft pose estimation, Proximity operations,**

ACRONYMS/ABBREVIATIONS

Active Debris Removal (ADR)
Binary Robust Invariant Scalable Keypoints (BRISK)
Convolutional Neural Network (CNN)
Final Approach (FA)
Guidance, Navigation and Control (GNC)
Hyperparamter Optimization (HPO)
On-Orbit Servicing missions (OOS)
Oriented FAST an Rotated BRIEF (ORB)
Pose EStimation Network (PEN)
Perspective-n-Point (PnP)
Scale-Invariant Feature Transform (SIFT)
Space Rendezvous Laboratory (SLAB)
Spacecraft Pose EStimation Dataset (SPEED)
Speeded-Up Robust Features (SURF)
Tree-structured Parzen Estimator (TPE)
Testbed for Rendezvous and Optical Navigation (TRON)
Waking Safety Ellipse (WSE)

This work was carried out within the Space It Up! project funded by the Italian Space Agency (ASI) and the Italian Ministry of University and Research (MUR) under contract No. 2024-5-E.0, CUP No. I53D24000060005.

I. INTRODUCTION

In recent years, the proliferation of satellite launches has significantly increased the need for advanced guidance, navigation, and control systems to support proximity operations for a variety of applications, such as On-Orbit servicing (OOS) [1],remote sensing [2] and Active Debris Removal (ADR) [3]. These mission profiles require precise and robust pose estimation of non-cooperative targets to ensure safe rendezvous and docking.

Traditional vision-based pose estimation techniques typically rely on image registration through feature detection and matching algorithms. Comparative studies [4] have highlighted the performance trade-offs among various classical approaches. In particular, SIFT [5] and SURF [6] are known for their high accuracy and robustness; however, their computational complexity makes them less suitable for real-time or resource-constrained applications, such as those encountered in space missions. Binary descriptors such as ORB [7] and BRISK [8] offer significantly faster execution and reduced memory consumption, but often at the expense of matching accuracy and repeatability, particularly under challenging conditions such as extreme illumination changes and limited visual texture. Recently, KAZE [9] and Accelerated KAZE [10] attempted to balance this trade-off by leveraging nonlinear scale spaces to improve performance in low-texture regions, an advantageous property for space imagery where surface features can be sparse or homogeneous. Despite significant progress, many existing methods continue to face limitations when applied to the harsh and dynamic conditions characteristic of the space environment. These challenges have driven increasing interest in deep learning-based approaches, which aim to improve generalization and robustness for accurate pose estimation. However, early deep learning models are often

overparameterized and computationally demanding, rendering them impractical for deployment on the resource-constrained hardware typically available aboard spacecraft [11]. Moreover, the training datasets used in many studies lack sufficient variability—often featuring static trajectories, limited viewpoint diversity, and minimal background complexity—which can lead to overfitting and poor generalization to real-world orbital scenarios.

This work presents a novel CNN-based architecture for monocular pose estimation of non-cooperative spacecraft, specifically designed to predict both the relative position and orientation of a known but non-cooperative target. The network is trained on a high-fidelity synthetic dataset generated using Blender, encompassing approximately 25,000 labeled images that simulate realistic proximity operations under varying lighting conditions, backgrounds, and spacecraft poses. Emphasis is placed on enhancing the model's robustness and generalization by diversifying the training scenarios and carefully optimizing the network architecture. To assess the effectiveness of the proposed approach, an extensive benchmarking analysis is performed against several state-of-the-art methods using standardized metrics. The results demonstrate competitive performance in terms of accuracy and robustness, positioning this method as a promising solution for visual-based navigation in on-orbit servicing and active debris removal missions.

The paper is structured as follows: Section 2 presents and discusses the proposed network, focusing on its design rationale and training strategy. Section 3 includes a comparative analysis in which the proposed method is quantitatively evaluated against representative state-of-the-art techniques using key performance indicators to assess improvements in estimation accuracy and robustness under realistic operational scenarios. Finally, Section 4 concludes the paper by summarizing the main contributions and findings, outlining possible directions for future work and integration in real-world applications.

II. Methodology

A. Dataset

Synthetic images are used to train the network. In particular, two different scenarios have been taken into account: the Walking Safety Ellipse (WSE) for target observation, in which the target is placed in a random position and orientation with respect to the chaser, in a range that spans from 600 to 100 m of distance, and Final Approach (FA) trajectory, which represents a straight-line trajectory to be followed by the chaser to perform rendezvous and docking with the target along the InTrack direction, in a range that spans from 80 to 8 m. The dataset is generated with Python APIs for Blender [12] and it is composed of 20000 images for WSE and 5000 for FA. Data augmentation is also performed to enhance the performances and reduce the overfitting on synthetic images. The applied effects can be divided into style augmentation, like *Blur* or *Texture Randomization* and more standard ones, like *Flip* and *Resize*.

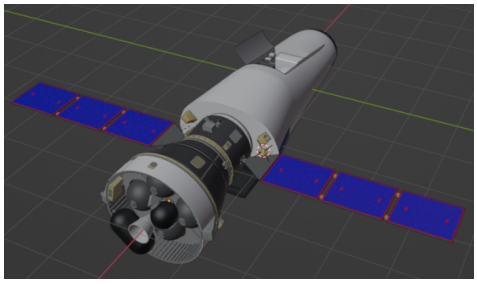

Fig. 1. Blender representation of the target

B. Network architecture

The proposed architecture follows a structured three-step process designed to robustly estimate the pose of a known, non-cooperative target spacecraft:

- **Keypoint Definition and Ground Truth Generation:** a set of 6 semantically meaningful keypoints is selected on the spacecraft model, including the head, tail, and the endpoints of the solar panels. The corresponding 3D coordinates are retrieved from the CAD model of the target, and the ground truth poses are computed by solving the inverse PnP problem [13].

- **Image Filtering via Binary Classification:** a binary classifier is used to detect the presence of the target in each image. This step filters out irrelevant samples, such as black images or those captured in eclipse conditions, where no meaningful visual features can be extracted.

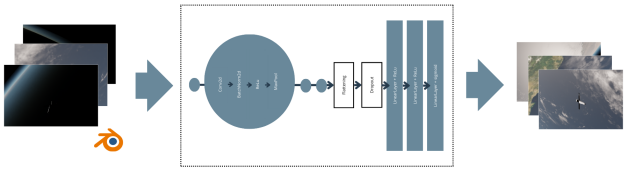

Fig. 2. Binary classifier

- **Multi-Head Prediction Network**
  The filtered images are then processed by a second CNN composed of multiple prediction heads. Each head is specialized for a specific task: bounding box detection, keypoint localization, and direct pose regression.

All the heads are connected by the Shared Feature Encoder composed by the EfficientNet [14] backbone and the Bi-directional Feature Pyramid Network [15]. The presence of multiple heads has the aim to collect more generalized features that can add robustness to the prediction.

III. Results and Discussion

A. Performance of the proposed method

The first CNN is trained and evaluated on both datasets using a 70-15-15 split for training, validation, and testing, respectively. Its primary objective is to classify each image based

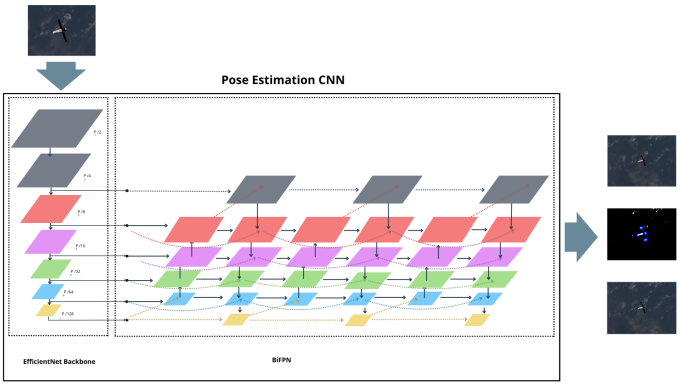

Fig. 3. Pose Estimation Network

on the presence or absence of the target spacecraft, assigning labels of Not Present (0) or Present (1). The generated dataset includes a variety of scenarios, such as images containing the spacecraft, fully black frames, and eclipse scenes in which the target is barely visible, making the classification task particularly challenging. To support this filtering process, a confidence threshold was introduced, initially set at 0.25 and later increased to 0.50, since the model demonstrated great ability in accurately detecting the target even under poor lighting conditions. The network achieves a test accuracy of 97.42%, highlighting its robust discriminative capability. This initial classification step is essential for automatically removing irrelevant or low-quality images, thereby significantly enhancing the quality of the dataset used for downstream pose estimation.

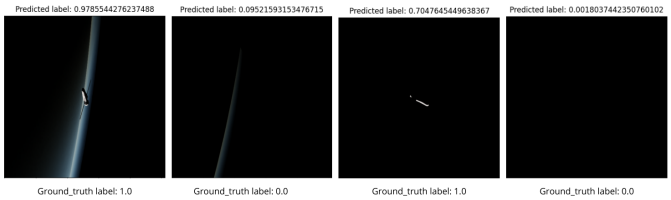

Fig. 4. Examples of image classifications in different conditions. Particularly, the first image presents a visible target, the second image presents a totally eclipsed target, in the third image the target is partially eclipsed but still recognizable, while in the fourth image no target is present

The second CNN is trained from scratch on both cleaned datasets using the same 70-15-15 split. The input images, prefiltered by the binary classifier, are resized from 2048×1536 to 1024×768 pixels to reduce computational load while preserving aspect ratio and prediction accuracy. Training is performed over 100 epochs (approximately 196 hours or 8.17 days) on an NVIDIA GeForce RTX 3090 GPU, resulting in a model specifically tailored to the task without relying on pre-trained features. The network includes two prediction heads—Heatmaps and EfficientPose—whose performance varies depending on the operating conditions. A combined approach that leverages both heads is adopted for final pose estimation. The following section presents and discusses the evaluation metrics used to assess the network's performance.

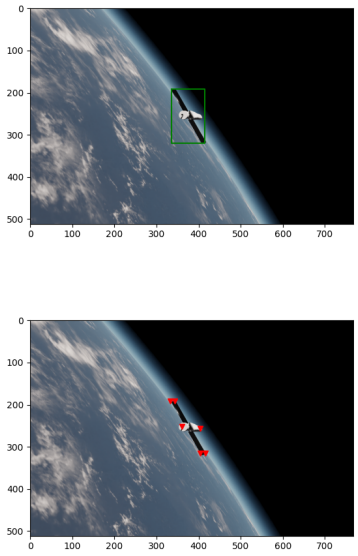

Fig. 5. Example of WSE dataset target detection and pose estimation through EfficientPose

SPEED loss is the official performance metric of the Satellite Pose Estimation (SPEC) challenge [16].It is designed to evaluate the predicted satellite pose in terms of both position and orientation. The position error e_t is calculated as the 2-norm differnce between the ground truth t_{gt} and the predicted position vectors t_{pr}.

$$e_t = \|t_{\text{gt}} - t_{\text{pr}}\|_2 \tag{1}$$

The normalized position error $\bar{e}_t$ is instead computed as

$$\bar{e}_t = \frac{e_t}{\|t_{\text{gt}}\|_2} \tag{2}$$

penalizing the position errors more heavily when the target satellite is closer. The orientation error e_q is computed as the angular distance between the true and predicted quaternions q_{gt} and q_{pr}, i.e., the magnitude of the rotation that aligns the the target body reference frame and the camera one.

$$e_q = 2arccos|q_{gt} \cdot q_{pr}| \tag{3}$$

The pose error e_{pose} is then calculated as the sum of the normalized position error and the orientation error.

$$e_{\text{pose}} = \bar{e}_t + e_q \tag{4}$$

The total error is then computed as the mean of the pose errors over the entire dataset.

$$e_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N} e_{\text{pose}_i} \tag{5}$$

TABLE I
WSE MEAN VALUE POSE ESTIMATION RESULTS EXPLOITING DIFFERENT CONFIGURATIONS. TABLE INSPIRED BY [16]

Synthetic dataset + Data Augmentation					
Config	**#Rej**	**IoU [-]**	E_t **[m]**	E_q **[deg]**	$E_{\textbf{pose}}$ **[-]**
E	-	0.935	0.724	4.602	0.094
H	1	-	0.779	5.215	0.118
E + H	1	0.935	**0.418**	**2.386**	**0.050**

TABLE II
FA MEAN VALUE POSE ESTIMATION RESULTS EXPLOITING DIFFERENT CONFIGURATIONS. TABLE INSPIRED BY [16]

Synthetic dataset + Data Augmentation					
Config	**#Rej**	**IoU [-]**	E_t **[m]**	E_q **[deg]**	$E_{\textbf{pose}}$ **[-]**
E	-	0.992	0.217	0.035	0.005
H	1	-	3.693	19.63	0.402
E + H	1	0.992	**0.213**	**0.035**	0.005

The table above shows the results of both datasets through the Pose Estimation Network using single-head/multiple-head approaches. The first column represents the configuration of the network (E stands for EfficientPose while H stands for Heatmaps), the second contains the number of rejected images (this feature is implemented in Heatmaps head since it can happen that noisy keypoints predictions result in outliers after pose regression), the third column shows the bounding box predictions, while the fourth and fifth columns show respectively the translation and the rotation error. The overall loss is at last presented in the sixth column. The results demonstrate that the network is capable of accurately estimating the spacecraft's pose with a high degree of precision and robustness. In particular, the most reliable predictions are obtained when the **multiple prediction heads** are employed **jointly**, highlighting the effectiveness of the multi-task learning approach.

The results of the two different datasets are compared in the table below. The only difference lies in the values of loss

TABLE III
PERFORMANCE COMPARISON OF PEN PERFORMANCES BETWEEN WSE DATASET (FIRST ROW) AND FINAL APPROACH DATASET(SECOND ROW)

Synthetic dataset + Data Augmentation					
Resolution	**#Rej**	**IoU [-]**	E_t **[m]**	E_q **[deg]**	$E_{\textbf{pose}}$ **[-]**
1024x768	1	0.935	0.418	2.386	0.085
1024x768	1	0.992	0.213	0.035	0.005

heads (0.096 VS 0.011 for pose loss). This can be attributed to many reasons: firstly, the final approach dataset is trained using pre-trained weights from the general dataset, allowing it to learn faster. Secondly, while the resolution is the same, the final approach dataset has a less cluttered scene (the Earth is barely visible in the beackground) with adjustments made to create a more photorealistic dataset. Also, following a straight-line trajectory, the pose of the final approach dataset is *easier* to learn.

B. Benchmarking and comparative analysis

Several datasets and architectures are analyzed to be compared with the one developed in this work (highlighted in gray in the tables). The considered datasets are (in order of appearance in the table below):

- **COSMO Photorealistic Dataset (first row):** This dataset consists of a collection of 15000 RGB images, each with a resolution of 1900x1200 pixels. The images are produced using a satellite from the SkyMed Earth observation constellation, placed in various poses, solar array configurations, lighting conditions, and backgrounds. In each image, the Earth is visible, and a sun lamp is used to simulate the Sun. The distance is randomly taken from a standard normal distribution with a mean of 36 m and a variance of 10 m, excluding all values under 36m or over 70m.*"The x-y offsets on the image plane are chosen randomly from a multivariate normal distribution, and they are constrained to ensure that the satellite is almost always entirely in the image frame. The attitude of the satellite is randomly selected from a uniform distribution of rotations. The images are post-processed to include a glare node (to replicate the bloom effect), greyscale conversion, Gaussian noise, and Gaussian blurring to emulate shot noise and depth of field* [17]".
- **WSE dataset (second row seventh row):** it is revlevant to note that the dataset presented in the seventh row [13] is an early version of the one presented in this work, so it presents the same characteristics, but with a different number of images (5000 instead of 20000) and resolution used for training (2048x1536 instead of 1024x768). Also, cleaning from black and shadowed images is done manually by hand reducing the dataset to 4500 images.
- **FA dataset (third row)**
- **SPEED+ [18] (fourth and fifth row):** it is composed of 59660 B&W synthetic images of the Tango spacecraft to train the network and respectively 6740 HIL images called lightbox, a half-scale mockup model of the Tango spacecraft illuminated with albedo lightboxes to simulate diffuse light in Earth's orbit and 2791 HIL images called sunlamp, images of the same model illuminated with a metal halide arc lamp to simulate direct sunlight. The image's size is 768x512 pixels and the target is placed at a maximum of 10 m of distance. In the fourth row, the dataset is considered as it is, while in the fifth row (*) a portion of it of the same amount as the WSE dataset is used to validate the Pose Estimation Network built in this work, so the 70-15-15 ratio is applied to it. In this dataset, no scene is present behind the target since the images are taken from the first 30000 pictures of the dataset to test the performances of the network only focusing on the target. Another possible trial could be to use a portion of the dataset where also background is

present to quantify the disturbance of the background on the network's performance.

- **SPEED [19] (sixth row):** it is composed of 15000 synthetic images of the Tango spacecraft. It is composed of 15000 B&W synthetic images of the Tango spacecraft from the Prisma mission [20] and 300 HIL images captured in *Testbed for Rendezvous and Optical Navigation* (TRON) facility at SLAB. Also in this case, the used resolution is 768x512. Images are placed in random positions and at a range between 3 and 40 m of distance.

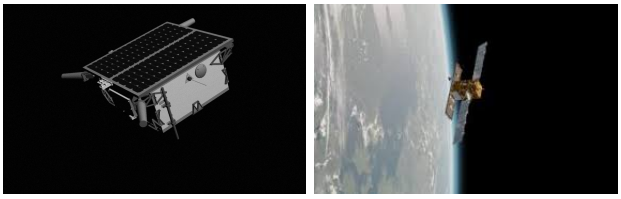

Fig. 6. Examples from COSMO Photorealistic Dataset (RGB) and SPEED+ Dataset (B&W) without background

Table IV is structured as follows: the first column lists the initial number of images available for each dataset, while the second column specifies their original resolution (noting that most datasets are downscaled prior to training). The next three columns indicate the number of training, validation, and test images, where applicable. Following this, the table reports the range of distances represented in each dataset, and finally, the number of training epochs used for each network.

TABLE IV
DIFFERENCES BETWEEN DATASETS USED IN OTHER NETWORKS AND THE ONES PRESENTED IN THIS WORK

Init. imgs	Dim.[pxl]	Pol. imgs	Train	Val.	Test	Dist.[m]
15000	1920x1200	-	15000	-	-	36 to 70
15000	2048x1536	9480	6636	1422	1422	100 to 600
5000	2048x1536	3014	2108	453	453	8 to 80
59660	1920x1200	59660	-	-	9531	max 10
59660	1920x1200	9480*	6636	1422	1422	max 10
15000	1920x1200	12000	-	2998	300	3 to 40
5000	2048x1536	4500**	4500	-	-	10 to 100

It is worth noting that the architectures of the evaluated networks differ significantly. For example, Lotti et al. [17] and Sharma et al. [19] adopt a two-stage approach comprising an object detection module—trained for 50,000 and 100 epochs, respectively—followed by a Keypoint Regression Network (KRN). In their setup, the target is first detected, and the corresponding region of interest (RoI) is cropped from the image to facilitate keypoint extraction. In contrast, the other networks considered in this study follow a single-stage design without such modular separation.

Additionally, differences in camera parameters used during dataset generation can influence pose estimation performance. The SPEED dataset is generated using a camera model with a 20mm focal length, while SPEED+ employed a 17mm focal length. The WSE dataset used a 70mm focal length, whereas the dataset proposed in this work simulates a camera with a 3mm focal length. Since focal length affects image perspective and scale, these variations can directly impact the difficulty of the estimation task. A comparative overview of the main metrics across these networks is provided in Table V

TABLE V
PERFORMANCE COMPARISON WITH STATE OF THE ART NETWORKS

Dim.[pxl]	Params	IoU [-]	E_{TN} [m]	E_q [deg]	E_{pose} [-]	Epochs
280x280	**10.5M**	0.968	**0.003**	0.55	0.0124	50000+450
1024x768	12.0M	0.935	0.043	2.386	0.085	100
1024x768	12.0M	**0.992**	0.004	**0.035**	**0.005**	100.pre
768x512	12.0M	> 0.919	0.009	1.224	0.031	20
512x384	12.0M	0.961	0.012	2.677	0.059	100.pre
224x1224	11.2M	0.919	0.019	3.097	0.073	100+300
2048x1536	59.0M	-	0.059	4.302	0.134	200

Various metrics are used to compare networks, such as image resolution, number of parameters, Intersection over Union (IoU), position error normalized E_{TN}, angle error in degrees E_q, pose error E_{pose}, and number of epochs. The network designed by Lotti et al. [17] has a reduced resolution of 280x280 pixels and 10.5 million parameters, making it ideal for real-time applications (a lightweight design is crucial given the limited computational resources of on-board devices). In comparison, all other analyzed networks have around 12 milions parameters, except for the ResNet [13] (seventh row), which exceeds this threshold, making the net too computationally heavy. In terms of detection performance, the model presented in this work shows the maximum accuracy of 99% when tested on the final approach dataset, whereas the other networks provide results of about 90%. The SPEED loss is used for pose estimation evaluation. The pose error is obtained from the sum of normalized distance errors and rotation errors (here reported in degrees). As indicated in the table, the Pose Estimation Network (PEN) offers the lowest pose loss at 0.005. The models used by Sharma et al. [19] and Lotti et al. [17] follow closely with 0.031 and 0.0124. It's also worth noting that despite undergoing fewer training iterations than the presented network (20 iterations versus 100), PEN achieves better performances despite the significantly larger SPEED+ dataset used in SPNV2 (it is approximately twelve times larger than the one used in the final approach configuration).

In conclusion, the network trained with the FA dataset achieved both high detection accuracy (IoU 0.992) and high score in position and pose estimation while keeping the number of parameters reasonable and the number of training epochs relatively low. It also demonstrated robustness across different metrics, indicating that the model could be suitable for real-time applications, considering both accuracy and computational efficiency. It is essential to note though that the used dataset comprises images that are not very cluttered since the Earth is barely visible in the background, and the pose does not vary much due to the straight-line trajectory of this maneuver. This can highly influence the performance of the network. The table also shows that processing larger images does not

always result in proportionally better performance, as might be expected.

CONCLUSIONS

This work introduces a CNN architecture designed to improve the pose estimation accuracy of known, non-cooperative spacecraft by merging proven methodologies with scenario-specific innovations. A synthetic dataset generated via Blender simulates realistic space mission environments, augmented with randomized lighting, textures, and backgrounds to enhance feature generalization. Experimental results demonstrate the model's centimeter-level positional accuracy and near-degree attitude precision, alongside robustness to illumination variations and cluttered space environment (e.g. the presence of ther Earth in background adds many different textures to the image). The study further includes a preliminary model comparison analysis, evaluating performance against state-of-the-art architectures to validate design choices and quantify trade-offs between computational efficiency and pose estimation fidelity. This analysis aims to identify optimal configurations for resource-constrained platforms while maintaining mission-critical accuracy. While promising, the results presented are based on synthetic data, which, though diverse, lacks the real-world complexity and conditions of actual space missions,with a possibility of overestimating the model's actual effectiveness in real mission scenarios.Future work will focus on addressing these limitations by building real-world datasets and further validating the model's performance in more complex and varied space environments. Additionally, efforts will be made to expand the dataset diversity to include more complex scenarios and improve model robustness in realistic conditions. The integration of segmentation modules for target isolation, along with refining model lightweighting strategies to balance precision with deployability, will be explored. Hardware-in-the-loop testing will assess real-time performance on flight-grade embedded systems, guiding further optimization for autonomous rendezvous, proximity operations, and on-orbit servicing missions.

REFERENCES

[1] S. Corpino et al., "Space rider observer cube - sroc: A cubesat mission for proximity operations demon- stration," in Proceedings of the 73rd International Astronautical Congress, IAC 2022, Code 190266, Paris, 2022.

[2] N. Schwartz et al., "6u cubesat deployable tele- scope for optical earth observation and astronomical optical imaging," in Proc. SPIE 12180, Space Telescopes and Instrumentation 2022, 2022. DOI: 10 . 1117 / 12 . 2627248.

[3] Robin Biesbroek, Sarmad Aziz, Andrew Wolahan, Stefano Cipolla, Muriel Richard-Noca, and Luc Piguet. The ClearSpace-1 Mission: ESA and ClearSpace Team Up to Remove Debris. In: Proceedings of the 8th European Conference on Space Debris. Darmstadt, Germany: ESA Space Debris Office, Apr. 2021. url: http://conference.sdo.esoc.esa.int

[4] R. H. Aravind and K. R. Santhosh, A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK, in *Materials Today: Proceedings*, vol. 52, Part 1, 2022, pp. 1510–1516. doi: 10.1016/j.matpr.2021.11.373.

[5] D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2, doi: 10.1109/ICCV.1999.790410.

[6] Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded up robust features. European Conference on Computer Vision (2006)

[7] E. Rublee et al., "ORB: An efficient alternative to SIFT or SURF," in IEEE International Conference on Computer Vision, Barcelona, ICCV,2011, pp. 2564-2571.

[8] S. Leutenegger et al., "BRISK: Binary robust invariant scalable keypoints," in IEEE International Conference on Computer Vision, Barcelona, ICCV, 2011, pp. 2548-2555.

[9] P. F. Alcantarilla et al., "KAZE features," in European Conference on Computer Vision, Berlin, ECCV, 2012, pp. 214-227

[10] P. F. Alcantarilla et al., "Fast explicit diffusion for accelerated features in nonlinear scale spaces," in British Machine Vision Conference, Bristol, BMVC, 2013.

[11] Silvia Rea, Chiara Tortora, Marco Zoppi, and Lorenzo Moriconi. A Survey of Deep Learning-Based Pose Estimation Methods for Non-Cooperative Spacecraft Rendezvous. In: arXiv preprint arXiv:2305.07348 (2023). url: https://arxiv.org/abs/2305.07348 (cit. on pp. XX).

[12] Lucrezia Lovaglio, Antonio D'Ortona, Fabrizio Stesina, and Sabrina Corpino. CNN-Based Visual Navigation: Optimization Strategies for Monocular Pose Estimation in Proximity Operations. In: *Proceedings of the IAF Astrodynamics Symposium*. Milan, Italy, Oct. 2024, pp. 1470–1480. doi: https://doi.org/10.52202/078368-0127
****

[13] D'Ortona, G. Daddi. "Relative visual navigation based on CNN in a proximity operation space mission" In: Aerospace Science and Engineering, 2023

[14] Mingxing Tan and Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *arXiv preprint* arXiv:1905.11946 (2019). url: http://arxiv.org/abs/1905.11946

[15] Mingxing Tan, Ruoming Pang, and Quoc V. Le. EfficientDet: Scalable and Efficient Object Detection. In: *arXiv preprint* arXiv:1911.09070 (2019). url: http://arxiv.org/abs/1911.09070.

[16] Kisantal, Mate & Sharma, Sumant & Park, Tae & Izzo, Dario & Märtens, Marcus & D'Amico, Simone. (2020). Satellite Pose Estimation Challenge: Dataset, Competition Design and Results. IEEE Transactions on Aerospace and Electronic Systems. PP. 1-1. 10.1109/TAES.2020.2989063.

[17] Alessandro Lotti, Dario Modenini, Paolo Tortora, Massimiliano Saponara, and Maria A. Perino. Deep Learning for Real Time Satellite Pose Estimation on Low Power Edge TPU. 2022 (cit. on pp. 69, 71, 72).

[18] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. SPEED+: Next-Generation Dataset for Spacecraft Pose Estimation across Domain Gap. In: 2022 IEEE Aerospace Conference (AERO). 2022, pp. 1–15. doi: 10.1109/AERO53065.2022.9843439 (cit. on p. 70).

[19] Mate Kisantal, Sumant Sharma, Tae Ha Park, Dario Izzo, Marcus Martens,and Simone D'Amico. Satellite Pose Estimation Challenge: Dataset, Competition Design, and Results. In: IEEE Transactions on Aerospace and Electronic Systems 56 (Oct. 2020), pp. 4083–4098. doi: 10.1109/taes.2020.2989063. url: http://dx.doi.org/10.1109/TAES.2020.2989063 (cit. on pp. 27, 70–72).

[20] D'Amico, Simone, John L. Christian, and Marcus Lavagna. The PRISMA Mission: An Overview and Lessons Learned. In: *Proceedings of the 4th International Conference on Spacecraft Formation Flying Missions and Technologies*. Saint-Hubert, Canada, May 2011.