

SPACECRAFT RECOGNITION LEVERAGING KNOWLEDGE OF SPACE ENVIRONMENT: SIMULATOR, DATASET, COMPETITION DESIGN AND ANALYSIS

Mohamed Adel Musallam¹, Vincent Gaudilliere¹, Enjie Ghorbel¹,
Kassem Al Ismaeil¹, Marcos Damian Perez², Michel Poucet², Djamila Aouada¹

¹ SnT, University of Luxembourg, ² LMO

ABSTRACT

SPARK represents the first edition of the *SPAcecraft Recognition leveraging Knowledge of space environment* competition organized by the Interdisciplinary Centre for Security, Reliability and Trust (SnT) in conjunction with the 2021 IEEE International Conference in Image Processing (ICIP 2021). By providing a unique synthetic dataset composed of 150k annotated multi-modal images, SPARK aims at encouraging researchers to develop innovative solutions for space target recognition and detection. This paper introduces the proposed dataset and provides a global analysis of the results obtained for the 17 submissions.

Index Terms— Dataset, Simulator, Space Situational Awareness (SSA), Object Recognition, Localization, Spacecraft, Space Debris.

1. INTRODUCTION

Today, our daily life is highly dependent on the increasingly growing satellite infrastructure. This infrastructure is used in many services; from communication, transportation, to weather forecast and many more. It is therefore essential to protect all space assets. One of the major threats is the risk of collision in space. Providing the spacecraft with the ability to autonomously recognize the surrounding objects is of utmost importance to minimize such risk and is one of the main objectives of Space Situational Awareness (SSA). Objects in question include active and inactive satellites, as well as space debris. Over the past years, image-based sensors have been considered as a great source of information for SSA. This has triggered multiple research efforts in the field [1, 2, 3, 4, 5], and recently, multiple works have been proposed to investigate the potential of deep learning (DL) from images for space applications [6, 7].

The performance of DL methods is highly dependent on the availability and quality of data used for training them. However, in the space domain, data are very scarce and costly

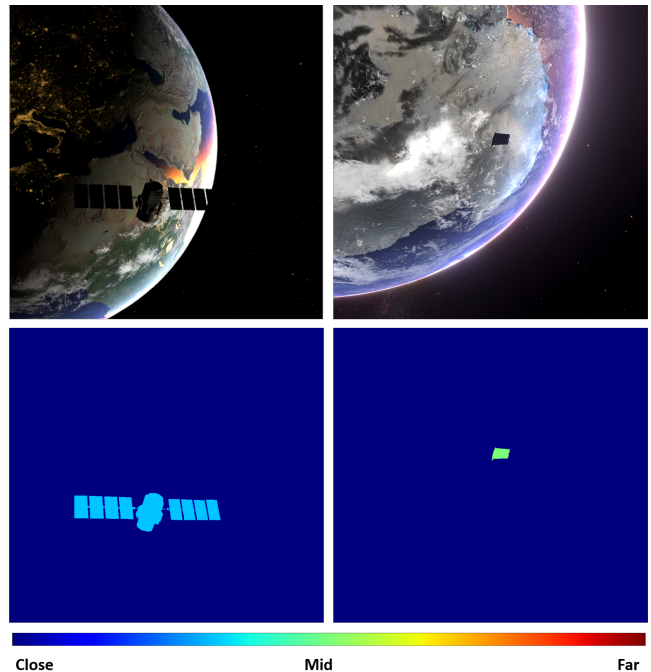


Fig. 1. Samples from our *SPARK* dataset. Top-left: RGB image of the ‘*Calipso*’ satellite with night side of Earth in the background. Top-right: RGB image of a debris with day side of Earth in the background. Bottom: corresponding depths.

to obtain. Recently, efforts have been initiated to propose four dedicated synthetic (or mixed with laboratory-acquired data) datasets, namely, the *Spacecraft pose estimation dataset (SPEED)* [8] [9], the *Unreal Rendered Spacecraft On-Orbit (URSO)* dataset [7], the *Swiss Cube* dataset [10] and the one published in [11], to which we will refer as the *Spacecraft-Parts* dataset. While the three first ones have been designed for the purpose of 6D pose estimation, the last one is dedicated to the task of semantic segmentation. However, to the best of our knowledge, no data exist so far for the task of space target recognition and detection while the latter are extremely important for SSA and consequently crucial towards reaching autonomy in space. Moreover, in view of the major advances made in object recognition using DL [12], it

This work was funded by the Luxembourg National Research Fund (FNR), under the project reference BRIDGES2020/IS/14755859/MEET-A/Aouada, and by LMO (<https://www.lmo.space>).

is interesting to investigate the applicability of these methods for space data, and identify directions for furthering their performance in the space environment.

Hence, the SPARK competition which provides a new public space annotated image dataset¹ dedicated for the task of space target recognition and detection comes just at the right time. Compared to the aforementioned datasets, the introduced benchmark is highly relevant since: 1) it is composed of two different modalities (RGB and depth images); 2) it contains a record-breaking number of 150k RGB images and 150k depth images (against a maximum of 21K for other datasets); and 3) it is formed by 11 object classes, with 10 spacecrafts and one class of space debris. Sample images are shown in Fig.1. Furthermore, it is important to note that the data have been generated under a photo-realistic space simulation environment, with a large diversity in sensing conditions, including extreme and challenging ones. It also incorporates other features already existing in previous datasets such as the 6D pose annotations. A detailed comparison of the SPARK dataset with existing benchmarks is shown in Table 1.

In this paper, the SPARK dataset and competition methodology are presented, while the submissions are analyzed and reported in aggregate form.

The rest of the paper is organized as follows: Section 2 describes the proposed SPARK dataset. The details related to the challenge are given in Section 3 including the analysis of the submissions. Section 4 concludes this work.

2. DATASET

The proposed SPARK dataset was generated using the Unity3D game engine as a simulation environment [13]. The simulation consists of the following main components:

Earth model: It consists of a high resolution textured and realistic Earth model composed of 16k polygons, obtained from a third party Unity asset [14], based on the NASA Blue Marble collection [15]. It presents clouds, clouds' shadows, and atmospheric outer scattering effects. This model is located at the center of the 3D environment with respect to all orbital scenarios. The surrounding space background uses a high resolution panoramic photo of the Milky way galaxy from the European Southern Observatory (ESO)[16].

Target spacecraft: Ten different realistic models of spacecrafts were used ('AcrimSat', 'Aquarius', 'Aura', 'Calipso', 'CloudSat', 'Jason', 'Terra', 'TRMM', 'Sentinel-6', and the 'IRU Generic CubeSat'). They were obtained from NASA 3D resources [17]. The debris are parts of satellites and rockets after adding corrupted texture to simulate dysfunction conditions ('space shuttle external tank', 'orbital docking system', 'damaged communication dish', 'thermal protection tiles', and 'connector ring'). They were placed around the

Earth and within the low Earth orbit (LEO) altitude.

Chaser spacecraft: It represents the observer equipped with different vision sensors used to acquire data.

Camera: A pinhole camera model was used with known intrinsic camera parameters and optical sensor specifications, as well as a depth camera for range imaging.

The SPARK dataset is generated by randomly placing the target satellite or spacecraft within the field of view of a camera mounted on a chaser. We consider different ranges and orientations of the chaser model. Furthermore, the Sun and Earth are randomly rotated around their respective axes in every frame. This allows to obtain high-resolution photo-realistic RGB images, depth maps, and the corresponding segmentation masks in multiple and different environmental conditions. The quality of spaceborne imaging is highly dependent on many factors such as varying illumination conditions, signal-to-noise ratio, and high contrast. Therefore, the SPARK dataset has been created in a way that covers a wide range of cases, including extreme and challenging ones. The whole dataset is spanned by sampling the following axes:

Scene illumination: We model the prominence of the Sun flares, rays, and reflections on Earth from the space in order to represent the different illumination conditions and the high contrast of the spaceborne images. Extreme illumination cases are considered where the sunlight directly faces the optical navigation sensors or reflects from the target surface or the Earth, hence causing lens flare and optical sensor blooming effect.

Scene background: In different orbital scenarios, the target spacecraft can be oriented towards the Earth or towards a dark space, which means that the target spacecraft's background can change with respect to different positions and orientations. When the Earth is in the background, the scene will have additional features of the planet's surface and high reflectivity of oceans and clouds. Whereas, having a dark space in the background will lead to featureless space with sparsely illuminated stars in the background.

Distance between the camera and the target: We model different ranges between the target spacecraft and the optical sensor mounted on the chaser spacecraft. Range is inversely proportional to the percentage of target occupation of the scene.

Optical sensor noise: Spaceborne images suffer from high noise levels due to small sensor sizes and high dynamic range imaging [6]. Accordingly, varying levels of zero-mean white Gaussian noise were added to the generated synthetic images to simulate the noise in real spaceborne images.

The final SPARK dataset consists of $\sim 150k$ high-resolution photo-realistic RGB images with bounding box annotation for the target object in each image, $\sim 150k$ depth images, and the corresponding $\sim 150k$ segmentation masks in multiple and different space environmental conditions. It includes 10 different satellites, 12.5k images for each satellite, and 5 debris objects, with 5k images for each one, and all debris combined

¹<https://cvi2.uni.lu/about-spark-2021/>

Dataset	SPARK (Ours)	SPEED [8]	URSO [7]	Spacecraft-Parts [11]	SwissCube [10]
# images per modality	150 k	15 k	15 k	2.5 k	21 k
# Spacecrafts	15	1	2	-	1
Visible	✓	✓	✓	✓	✓
Color	✓	✗	✓	✓	✓
Depth	✓	✗	✗	✗	✗
Mask	✓	✗	✗	✓	✓
3D Model	✓	✗	✗	✗	✓
6D Pose	✓	✓	✓	✗	✓
Rendering	simulated	simulated + lab	simulated	artists' views	simulated

Table 1. Comparison of Space Situational Awareness datasets.

in one class.

3. CHALLENGE

To introduce the dataset to the community, we have organized a Challenge Session in conjunction with the 2021 IEEE International Conference on Image Processing².

3.1. Competition Design

The goal of this Challenge was to investigate the interest of multimodality to tackle two major perception tasks in Space Situational Awareness: target object recognition and in-image localization. Throughout the rest of this document, we will refer to the former as the *classification* task and to the latter as the *detection* task.

3.1.1. Experimental Setup

The data provided to the participants is composed of 150 k RGB images along with their corresponding simulated depth maps (see Fig. 2). They are splitted into 90 k samples for training, 30 k for validation and 30 k for testing. In total, 15 different space objects are featured in the images: 10 satellites and 5 space debris.

For the classification task, the 5 debris were gathered into one class (named *debris*), whereas each satellite is considered as a full class. The main goal, for the participants, is to correctly recognize the spacecraft featured in each pair of RGB+depth images. In addition to correctly recognizing the spacecrafts, the detection task consists in correctly localizing their bounding box. Note that only the ground-truth of the training and validation subsets is released. It is constituted of spacecraft class labels and rectangular bounding boxes (see Fig. 2).

3.1.2. Metrics

To evaluate the different submissions and rank the participants, two metrics were specifically designed.

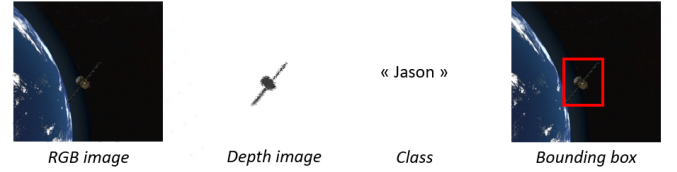


Fig. 2. Illustration of the Challenge data.

Task 1 - Classification

In this task, two types of errors can occur that are defined according to their level of seriousness as follows (from less to more serious): 1) classify debris as a satellite and, 2) Misclassify a satellite either as another satellite or as debris. Hence, the goal is to: 1) avoid confusing satellite and debris (with a higher penalization in the event of a misclassified debris); and 2) correctly classify the images featuring satellites based on their type. Based on these considerations, the following metric has been designed to rank the participants:

$$\text{Perf} = F_2\text{-score}(\text{debris}) + \text{accuracy}(\text{satellites}) \quad (1)$$

where

$$F_2\text{-score} = 5 \frac{\text{precision} \cdot \text{recall}}{4 \cdot \text{precision} + \text{recall}}$$

and accuracy is the proportion of correctly classified samples (averaged over the 10 satellite classes). Indeed, the F_2 -score is to consider recall (penalizing debris omissions) twice as important as precision (penalizing false debris detection). Adding the two terms allows giving the same importance for both 1) and 2) subproblems.

Task 2 - Detection

For this task, the localization accuracy is evaluated in addition to the classification performance. The metric is largely inspired by the one used in the COCO Challenge³. More precisely, it consists in computing the proportion of correctly predicted images in terms of classification for which the intersection-over-union (IoU) score between the predicted

²<https://www.2021.ieeeicip.org>

³<https://cocodataset.org/#detection-eval>

and the groundtruth bounding box is above a certain threshold. The final detection metric is then obtained by averaging these proportions over different IoU thresholds (between 0.5 and 0.95, with a step of 0.05).

3.2. Submission Analysis

In this section, we provide a global analysis of the SPARK challenge submissions.

A total of 17 submissions from 5 teams participated in the challenge. Each team provided a short report explaining their approach. We provide an analysis based on an overview of these approaches. However, note that we do not provide the ranking of participants, but rather focus on the overall performances.

The results obtained for the 17 submissions suggest that utilizing depth information may generally improve the performances for both tasks. The depth information has been used at two levels, either as a preprocessing step or by fusing it with the corresponding RGB images. As a preprocessing step, given that the object of interest can occupy a small region of the image, some teams used the provided depth images to first localize the satellite/debris, then to crop the relevant part of the image. On the other hand, other teams have fused the depth images along with the RGB images during the training phase.

Furthermore, data augmentation techniques were used by multiple teams to maximize the performance for both tasks, such as Horizontal and vertical flip, Gaussian noise and blur, CutOut, MixUp, CutMix and Mosaic. Generally, the participated teams have used various architectures of well-known pre-trained models, such as ResNets [18],[19], and YOLO[20].

For the **Classification** task, the overall confusion matrix showing the average submission results is presented in Fig. 3. The results show that between 54% and 81% of samples from each class were correctly classified, demonstrating a reasonable level of difficulty. One can also note that satellite misclassifications most often go for debris, justifying the special emphasis put on this class throughout the Challenge design process.

For the **Detection** task, the overall per-class detection results are shown with respect to the IoU threshold in Fig. 4. As expected, the obtained results show that, for most of the classes, increasing the IoU threshold leads to a decrease of the detection score and vice versa. We can also see that *CloudSat* is the most challenging object for both tasks (detection and classification). In order to understand this behavior, further investigation is required.

4. CONCLUSION

We proposed the SPARK dataset for spacecraft recognition, localization, segmentation and pose estimation. This dataset

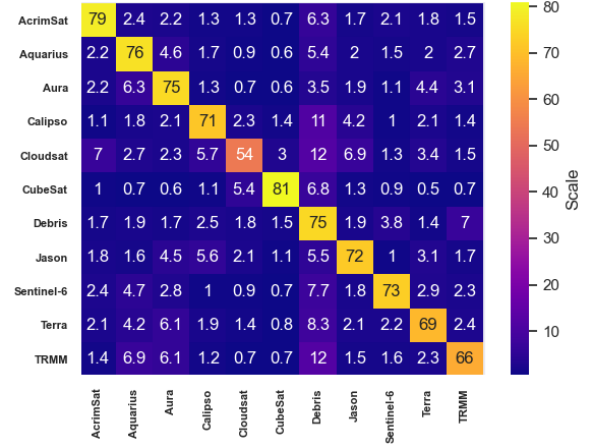


Fig. 3. SPARK Challenge 2021 (Task 1 – Classification): overall confusion matrix.

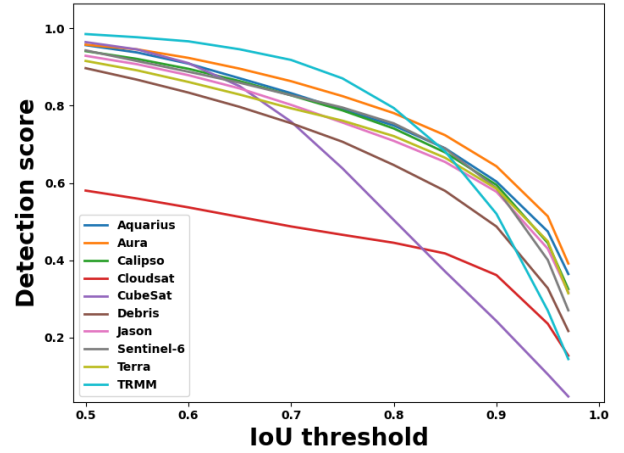


Fig. 4. SPARK Challenge 2021 (Task 2 - Detection): Proportion of correctly predicted images per class, as a function of the IoU threshold.

was generated under a realistic space simulation environment, providing a large range of diversity in sensing conditions and multiple spacecraft classes. To pave the way for developing multimodal deep learning approaches for debris and spacecraft recognition and detection in the context of SSA missions, we have organized the SPARK Challenge in conjunction with ICIP 2021. The interest of the community for this Challenge demonstrates the relevance of multi-modal RGB-Depth spacecraft perception as a research direction.

5. REFERENCES

- [1] R. Opromolla, G. Fasano, G. Rufino, and M. Grassi, “A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations,” *Progress in Aerospace Sciences*, vol. 93, 2017.

- [2] M. Strube, R. Henry, E. Skeleton, J. V. Eepoel, N. Gill, and R. McKenna, "Raven: An on-orbit relative navigation demonstration using international space station visiting vehicles," *Advances in the Astronautical Sciences Guidance, Navigation and Control*, vol. 154, 2015.
- [3] A. Yol, E. Marchand, F. Chaumette, K. Kanani, and T. Chabot, "Vision-based navigation in low earth orbit," in *i-SAIRAS'16*, 2016.
- [4] T. Chabot, K. Kanani, A. Pollini, F. Chaumette, E. Marchand, and J. Forshaw, "Vision-based navigation experiment onboard the removedebris mission," in *10th Int. ESA Conference on GNC Systems*, 2017.
- [5] J. L. Forshaw, G. S. Aglietti, N. Navarathinam, H. Kadhemi, T. Salmon, A. Pisseloup, E. Joffe, T. Chabot, I. Retat, R. Axthelm, et al., "Removedebris: An in-orbit active debris removal demonstration mission," *Acta Astronautica*, vol. 127, 2016.
- [6] S. Sharma, C. Beierle, and S. D'Amico, "Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks," in *2018 IEEE Aerospace Conference*, 2018.
- [7] P. F. Proença and Y. Gao, "Deep learning for spacecraft pose estimation from photorealistic rendering," in *ICRA*, 2020.
- [8] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. Martens, and S. D'Amico, "Satellite pose estimation challenge: Dataset, competition design and results," *IEEE Trans. On Aerospace and Electronic Systems*, 2020.
- [9] "Pose estimation challenge," <https://kelvins.esa.int/satellite-pose-estimation-challenge/>.
- [10] Yinlin Hu, Sebastien Speierer, Wenzel Jakob, Pascal Fua, and Mathieu Salzmann, "Wide-depth-range 6d object pose estimation in space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2021, pp. 15870–15879.
- [11] Hoang Anh Dung, Bo Chen, and Tat-Jun Chin, "A spacecraft dataset for detection, segmentation and parts recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2021, pp. 2012–2019.
- [12] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. journal of computer vision*, vol. 128, 2020.
- [13] "Unity3d," <https://unity.com/>.
- [14] "Planets earth asset," <https://assetstore.unity.com/packages/vfx/shaders/planets-earth-mars-and-the-moon-17593>.
- [15] "Nasa blue marble resources," <https://visibleearth.nasa.gov/collection/1484/blue-marble>.
- [16] "European southern observatory images," <https://www.eso.org/public/images/>.
- [17] "Nasa 3d resources," <https://nasa3d.arc.nasa.gov/>.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2016.
- [19] M. Tan, R. Pang, and Quoc V Le, "Efficientdet: Scalable and efficient object detection," in *IEEE/CVF conference on computer vision and pattern recognition*, 2020.
- [20] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, "You only look once: Unified, real-time object detection," 2016.