YOOLO: You Only Orbit and Look Once

Anonymous CVPR submission

Paper ID *****

Abstract

In the context of the SPARK 2026 *Challenge, a deep learning pipeline for the simultaneous classification, detection, and segmentation of spacecraft orbiting Earth is developed. The challenge provides a large-scale synthetic dataset of orbiting spacecrafts, including image-level class labels, bounding boxes, and pixel-wise segmentation masks for ten real satellite platforms. Participants are required to design compact models capable of jointly solving classification, detection, and segmentation tasks under strict efficiency constraints.*

In this work, YOOLO (You Only Orbit & Look Once) is introduced as a compact and efficient deep learning pipeline capable of addressing multiple computer vision tasks within a unified framework. A YOLO11S-SEG *model is trained on the full training set for 20 epochs, achieving an accuracy score of 0.938 on the test set. The results demonstrate that YOOLO effectively balances model compactness and predictive performance, delivering encouraging classification, detection, and segmentation outcomes while ensuring fast inference. This makes the proposed approach promising for resource-constrained space-domain applications.*

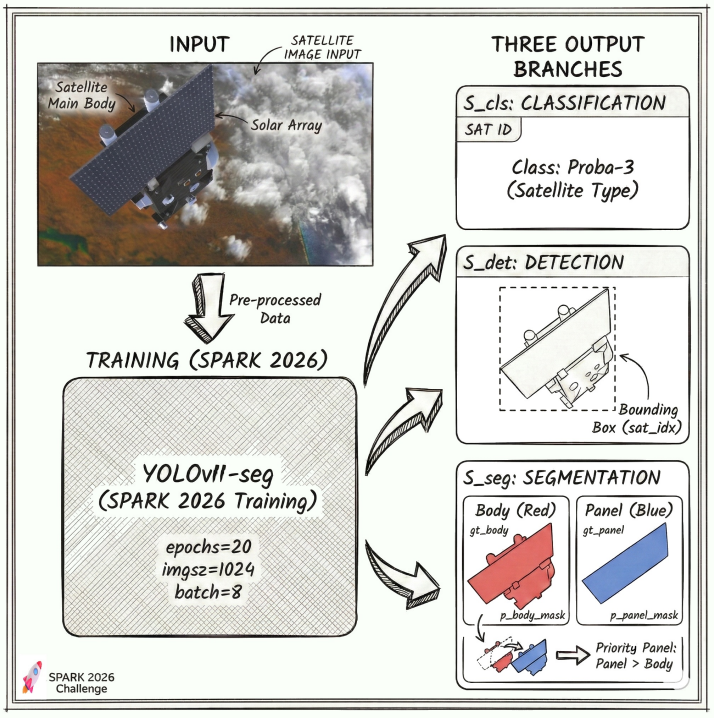

Figure 1. Sketch of the proposed model YOOLO (You Only Orbit and Look Once).

1. Introduction

Space Situational Awareness (SSA) focuses on the detection, monitoring, and characterisation of objects orbiting the Earth [1]. In recent years, the number of active spacecraft has increased rapidly, leading to severe congestion in orbital environments, particularly in Low Earth Orbit (LEO), and to the proliferation of space debris [5, 8], which could jeopardise the success of space operations. According to the ESA Space Environment Report 2025 [2], in 2025 space surveillance networks track approximately $40,000$ objects in Earth orbit, of which about $11,000$ are active satellites. Model estimates place the number of debris fragments larger than 1 cm at over 1.2 million, a size capable of causing catastrophic damage upon collision. This growing complexity makes SSA a critical research area for ensuring the safety and sustainability of space operations. Accurate and continuous monitoring is essential to support on-orbit activities such as rendezvous and docking, in-space refuelling, satellite servicing, and collision avoidance [14]. In this context, vision-based sensors play a key role, enabling autonomous perception and real-time decision-making through the use of Computer Vision (CV) techniques. Deep Learning (DL) has become the backbone of modern perception systems, achieving state-of-the-art performance in classification, detection, and segmentation tasks[4, 10]. However, DL models typically require large amounts of labelled data for effective training. In the space domain, the availability of annotated datasets remains limited, significantly constraining the deployment of

data-driven approaches. The SPARK 2026 Challenge addresses this limitation by providing a large-scale synthetic dataset of spacecraft images, generated from real orbital assets [12, 13, 15]. Participants were tasked with developing automated pipelines capable of simultaneously performing spacecraft classification, detection, and segmentation, while prioritising high accuracy and a low number of model parameters. Several recent works have explored YOLO-based and deep learning approaches for space and satellite imagery analysis. Extensions of the YOLO family have been proposed for spacecraft component detection and segmentation, such as the SpaceYOLOv2 [9] architecture that combines shape detection with contextual reasoning and improves recall over standard YOLOv5 classifiers in hardware-in-the-loop settings. Other algorithms like YOLO-Dynamic [16] introduce multi-scale feature fusion and lightweight neck designs to enhance small object recognition in space-borne imagery, outperforming baseline YOLOv8 models. Deep learning systems using YOLO variants have also shown promise for small space debris detection in radar maps and all-sky optical camera images, demonstrating the flexibility of YOLO frameworks in space domain scenarios [11]. Additionally, lightweight segmentation frameworks based on YOLO have been developed for spacecraft component segmentation, achieving high precision while maintaining efficiency. SLD-YOLO [6] further demonstrates state-of-the-art detection with attention enhancements on synthetic satellite image datasets. These studies collectively illustrate the relevance and adaptability of YOLO-based architectures for space object detection and segmentation tasks, motivating our adoption of a YOLO11-derived framework in the SPARK 2026 challenge.

In this paper, we introduce YOOLO, a compact multi-task deep learning pipeline designed to solve these three CV tasks within a unified framework. Building upon the YOLO11 architecture [3], a single, efficient model is proposed, optimised for low parameter count and fast inference, making it suitable for resource-constrained space applications. A sketch of YOOLO is presented in Fig. 1. The main contributions of this work are:

- YOOLO, a compact multi-task deep learning pipeline capable of performing spacecraft classification, detection, and segmentation in a unified model.
- Efficient architecture design optimized for low parameter count and fast inference, suitable for resource-constrained space platforms.
- Extensive evaluation on the SPARK 2026 dataset, demonstrating encouraging performance while maintaining an average computational footprint.

2. Dataset

Synthetic datasets represent a practical alternative to real imagery, partially mitigating the scarcity of large-scale, fully annotated datasets required for training deep learning models. Nevertheless, models trained exclusively on synthetic data often suffer performance degradation when deployed in real-world environments, primarily due to overfitting to synthetic feature distributions and the resulting domain gap. As discussed in [12], improving the physical realism of the data generation process is a key strategy to alleviate this issue. In particular, enhanced realism can be achieved by accurately reproducing solar-like illumination conditions, selecting materials with physically consistent reflectance properties, and employing physical spacecraft mock-ups within controlled laboratory environments to acquire representative imagery. The dataset released for the SPARK 2026 Challenge was provided by the SnT Zero-G Lab. It consists of a large-scale synthetic image collection comprising 10 spacecraft classes, generated from real orbiting platforms. The classes present in the dataset are respectively: VenusExpress, Cheops, LisaPathfinder, ObservationSat1, Proba2, Proba3, Proba3ocs, Smart1, Soho, XMM Newton. Each sample includes an RGB image, a class label, bounding-box annotations, and pixel-level segmentation masks. The masks first separate the spacecraft from the background and subsequently decompose it into its two principal structural components, namely the main body and the solar panel. The dataset contains 60,000 training images and 20,000 validation images, corresponding to 6,000 and 2,000 samples per class, respectively. Additionally, a test set of 5,000 unlabeled images is provided for evaluation. The dataset spans a broad spectrum of spacecraft types and mission profiles, ranging from Solar System exploration to platforms dedicated to extragalactic observation. This diversity ensures strong representational variability and reflects the intrinsic complexity of the classification task, closely approximating the heterogeneity encountered in real-world visual scenarios. The dataset exhibits substantial intra-class variability in terms of scale, orientation, illumination angle, and relative pose. Spacecraft instances appear under diverse sun–object–camera configurations, leading to pronounced shadowing and specular highlights. Additionally, variations in object scale and distance introduce challenges in small-object detection and boundary delineation. This controlled yet diverse distribution supports robust model training while preserving structured variability across samples.

3. Methodology

In the context of the YOLO family, a state-of-the-art prototype of end-to-end one-stage segmentation/classification model has been employed, YOLO11s-Seg [3], which has been compared to other all-in-one tasks solver for its semantic segmentation, image classification and object detection performance. Besides, YOLOv11-Seg is well-known for its lightweight structure, easy to mount on many kinds of

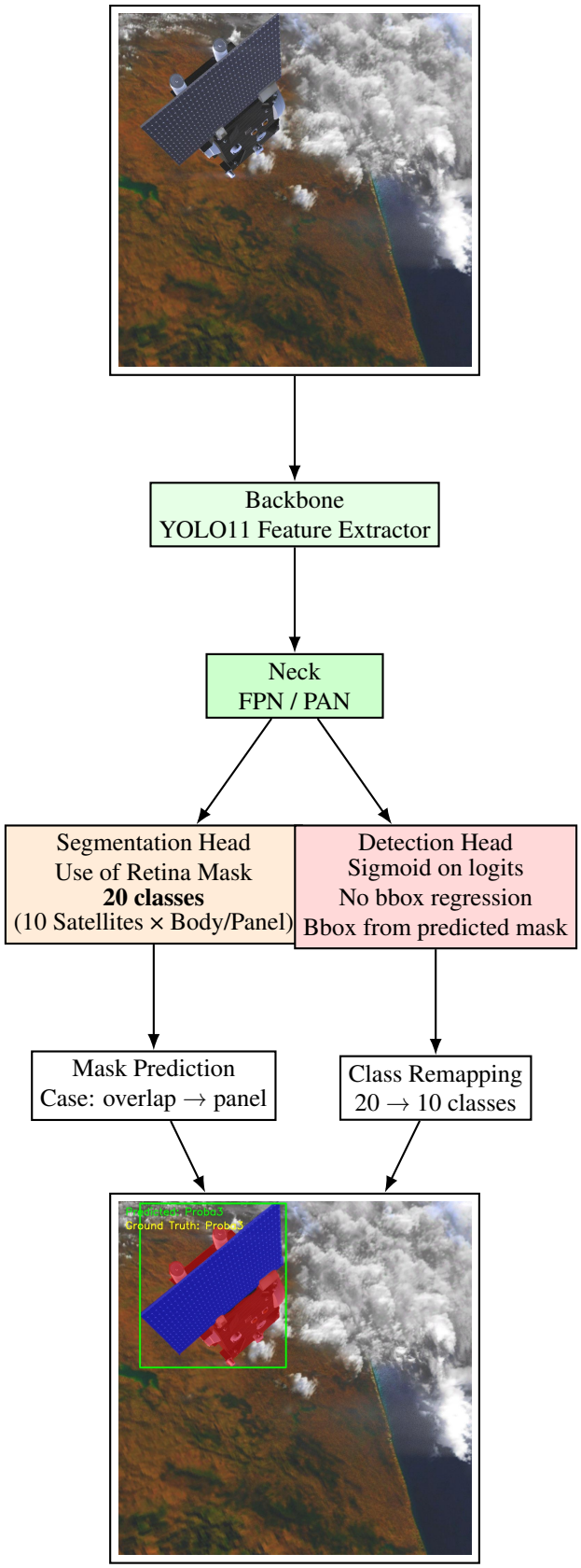

Figure 2. YOLOO (You Only Orbit and Look Once) architecture overview.

hardware and embedded GPU systems, so it was considered suitable for this problem. The overview of the architecture is shown in Fig. 2.

3.1. Baseline

With only 33.1 GFLOPs, the proposed method remains computationally competitive. The model has been pre-trained on COCO 330K with 80 classes [7] and consists of 355 layers, $10,090,028$ parameters, and $10,090,012$ gradients.

3.2. Classes encoding

The encoding of the SPARK Challenge 2026 data-set has been changed to be compatible with the YOLO11 dataset format. Every body and panel of each of the 10 spacecraft classes has been treated as a distinct class, so 20 possible combinations have been reported and annotated. The problem is remapped from 10 classes of objects to 20 classes during the training phase. During inference, the problem is projected again on the 10 classes, distinguished between panel and body for each detected spacecraft.

3.3. Architectural modifications

First, the model has been modified on the detection prediction branch: the standard detection branch bounding box has been noticed to be less effective on validation performance measurement, so the decisive prediction has been chosen as the minimum bounding box surrounding the two body and panel masks. This choice was crucial for the test set score assessment, as it will be shown in the Results and Discussion part. Second, the segmentation branch was modified to resolve overlaps between predicted body and panel pixels. Analysis on the validation set indicated that overlapping pixels predominantly belong to panels; therefore, a conditional rule was introduced to prioritize the panel class.

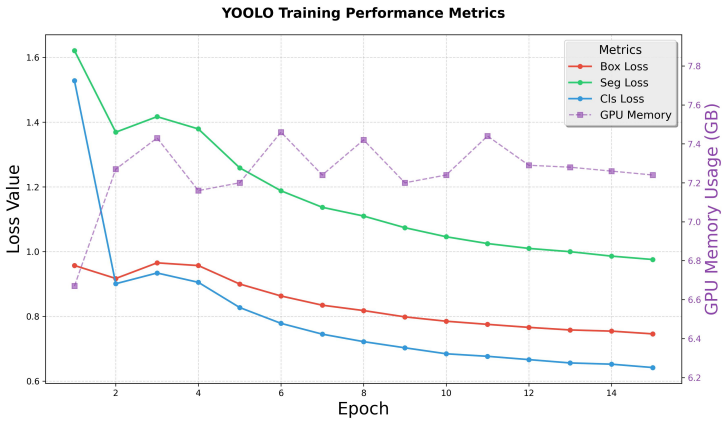

Figure 3. Training metrics obtained with YOOLO model on the 60K images SPARK Challenge 2026 training set.

Scoring functions The leaderboard score is computed as follows. The final score is

$$S_{\text{final}} = 0.7\, S_{\text{acc}} + 0.15\, S_{\text{flops}} + 0.15\, S_{\text{params}} \tag{1}$$

which balances accuracy (70%) with efficiency (30%). The multi-task accuracy is

$$S_{\text{acc}} = 0.2\, S_{\text{cls}} + 0.4\, S_{\text{det}} + 0.4\, S_{\text{seg}} \tag{2}$$

combining classification (20%), detection (40%), and segmentation (40%). Image-level classification is

$$S_{\text{cls}} = \frac{1}{N_{\text{img}}} \sum_{i=1}^{N_{\text{img}}} \mathbf{1}\Big\{\text{label of largest box}_i = \text{GT}_i\Big\} \tag{3}$$

measuring the fraction of images where the largest predicted object matches the ground truth. Object detection is

$$S_{\text{det}} = \frac{1}{N_{\text{GT}} T} \sum_{i=1}^{N_{\text{GT}}} \sum_{t=1}^{T} \mathbf{1}\{\text{IoU}(\text{pred}_i, \text{GT}_i) \geq \tau_t\} \tag{4}$$

computing mean success across all ground truth boxes and IoU thresholds ($\tau_t \in [0.5, 0.95]$, step 0.05). Segmentation is

$$S_{\text{seg}} = \frac{1}{|C| T N_{\text{valid}}} \sum_{c \in C} \sum_{t=1}^{T} \sum_{m \in \text{valid}} \mathbf{1}\{\text{IoU}(m, \text{GT}_c) \geq \tau_t\} \tag{5}$$

averaging IoU over Body and Panel masks that meet the 5% coverage rule. Finally, efficiency scores are

$$S_{\text{flops}} = \max(0, 1 - \text{FLOPS}/100) \tag{6}$$

$$S_{\text{params}} = \max(0, 1 - \text{Params}/30) \tag{7}$$

penalizing models exceeding 100 B-FLOPS or 30 M parameters.

Training set-up The model was trained for a total of **20** epochs, using early stopping based on validation performance, formalized in the S_{final} computation. Training was resumed from a previously saved checkpoint, and the best model was observed at epoch **16**. Input images were resized to **1024** $\times$ **1024** for both training and validation. Stochastic gradient descent (SGD) was used as the optimizer, with a learning rate of **0.01**, a momentum of **0.9**, and a weight decay of **0.0005**. Training and inference were performed on Python 3.12.10 with PyTorch 2.5.1+cu121 running on a NVIDIA GeForce RTX 4060 Laptop GPU with 8188 MiB VRAM. A chart with GPU vRAM consumption, bounding box loss, segmentation loss, and classification loss curves is depicted in Fig. 3, showing a successful optimization in the training phase with relatively low GPU memory usage. The YOLO confidence threshold for validation was set at **0.1**, chosen heuristically between **0** (too noisy) and **0.25**. Experimentally, it was observed that for the detection branch (after masks segmentation) a confidence threshold of **0.1** works fine, while for the segmentation branch the same happens with a confidence threshold of **0.25**.

Table 1. YOOLO validation results with ablation at confidence threshold 0.25.

Metric / Score	Conf 0.1	Conf 0.25
Model Summary		
Layers (fused)	265	
Parameters	10.07 M	
FLOPs	32.8 B	
Inference FPS	16.11	
Detection / Mask mAP		
Detection mAP@50	0.8874	–
Detection mAP@50-95	0.6778	0.6683
Mask mAP@50	0.9867	–
Mask mAP@50-95	0.8439	0.8520
Challenge Scores		
S_{cls}	0.9997	0.9994
S_{det}	0.6778	0.6683
S_{seg}	0.8497	0.8520
S_{acc}	0.8110	0.8080
$\mathbf{S_{final}}$	**0.7681**	0.7640

4. Results and Discussion

The performance of the YOOLO architecture was evaluated using the SPARK 2026 dataset. The validation phase was used to identify the optimal configuration via an ablation study, while the test set results provide a final measure of the model's accuracy.

4.1. Architectural Efficiency

The model consists of **265 fused layers** with a total of **10.07 M parameters** and requires **32.8 GFLOPs** for a single inference pass. Achieving a speed of **16.11 FPS**, the architecture represents a functional compromise for edge-computing applications in space scenarios. This efficiency allows for real-time monitoring of space assets even on hardware with limited computational power.

4.2. Model Selection (Validation)

Table 1 summarizes the trade-offs observed during the ablation study between two confidence thresholds.

The table shows that minimum confidence at **0.1** yields a higher detection performance with an mAP_{50} of **0.8874** and an mAP_{50-95} of **0.6778**. While increasing the threshold to **0.25** slightly improves the mask precision mAP_{50-95}

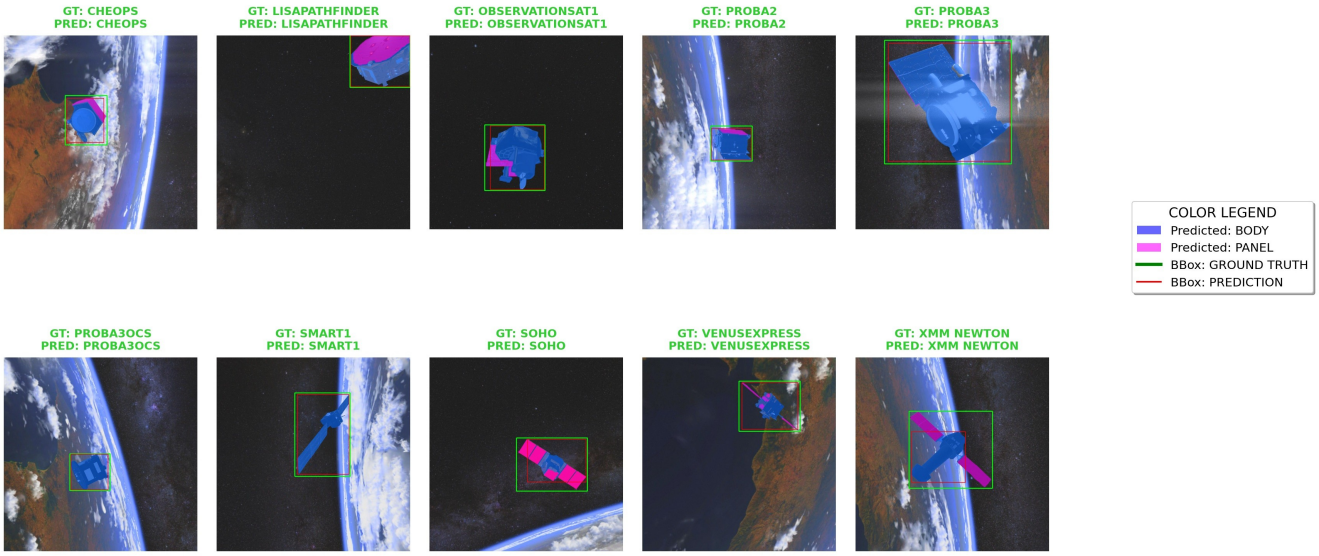

Figure 4. Visual results on SPARK 2026 Challenge validation set for each spacecraft class: in green the well predicted labels and ground-truth bounding boxes, in red the prediction bounding boxes, in cyan/blue the body prediction mask and in magenta the panel prediction mask.

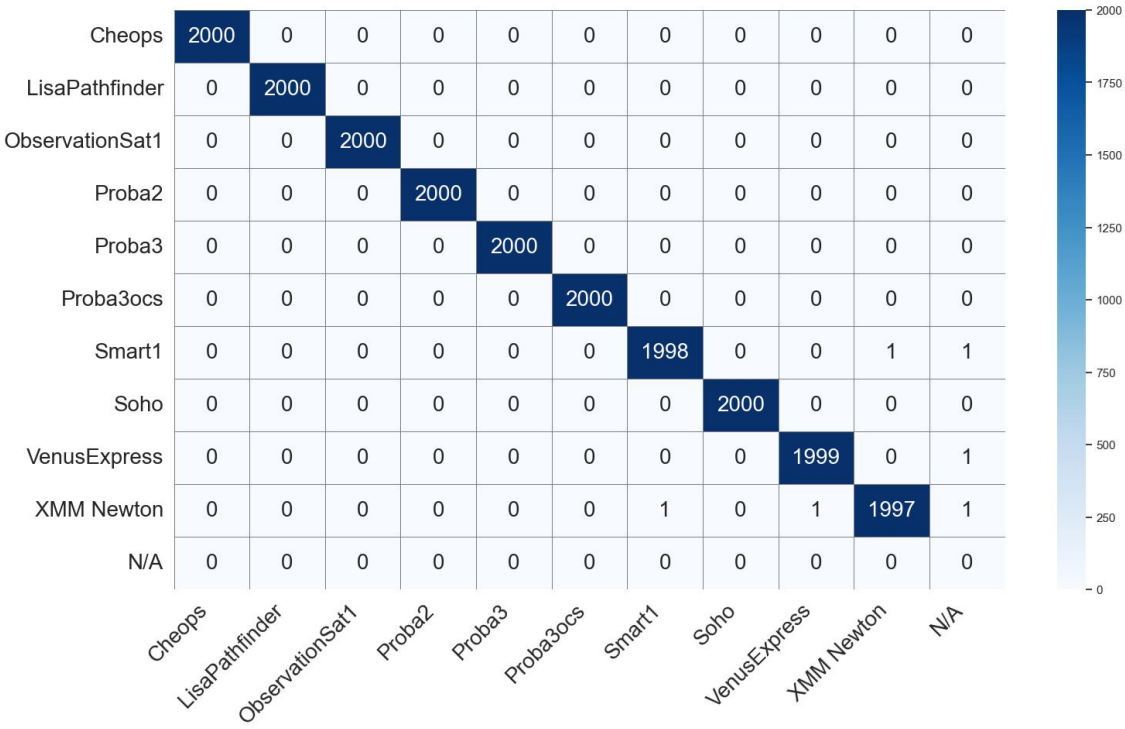

Figure 5. Confusion matrix of the classification experiments on SPARK Challenge 2026 validation set. N/A stands for "No Detection Occurred".

from **0.8439** to **0.8520**), it negatively impacts the classification S_{cls} drops from **0.9997** to **0.9994** and the detection recall. Consequently, the minimum-confidence **0.1** configuration was selected, as it maximises the final challenge score S_{final} at **0.7681**.

Table 2. Distribution of ambiguous pixels between body and panel components per satellite class in SPARK 2026 Challenge validation set.

Spacecraft	Body	Panel	Total	Body (%)	Panel (%)
Proba2	5,568,703	29,204,746	34,773,449	16.014	83.986
ObservationSat1	5,484,444	17,054,551	22,538,995	24.333	75.667
LisaPathfinder	1,652,191	14,518,214	16,170,405	10.217	89.783
XMM Newton	9,223,876	6,855,303	16,079,179	57.365	42.635
VenusExpress	8,128,572	6,958,913	15,087,485	53.876	46.124
Proba3ocs	2,538,724	7,950,741	10,489,465	24.203	75.797
Smart1	4,254,603	6,193,972	10,448,575	40.719	59.281
Cheops	1,371,129	6,262,551	7,633,680	17.962	82.038
Soho	1,994,163	4,047,191	6,041,354	33.009	66.991
Proba3	4,612,575	666,774	5,279,349	87.370	12.630

4.3. Ambiguities

Table 2 reports the distribution of ambiguous pixels between the body and panel components across satellite classes in validation set (i.e., pixels with mask prediction logits higher than **0.5** for both body and panel types, sharing the same value). A strong class-dependent imbalance in ambiguity attribution was observed. A very interesting distribution of logits was observed too: **0** where the model did not predict anything, **0.5** when the model was not secure between body and panel and **0.731** otherwise. Some satellites exhibit a pronounced dominance of panel ambiguity, such as **LisaPathfinder (89.783%)** and **Proba2 (83.986%)**, indicating that ambiguous regions are predominantly associated with solar panel structures. Similarly, **Cheops (82.038%)** and **ObservationSat1 (75.667%)** show a clear bias toward panel-related ambiguity.

Conversely, **Proba3** presents the opposite behavior, with **87.370%** of ambiguous pixels attributed to the Body component, suggesting a fundamentally different geometric or structural configuration. Other satellites such as **XMM Newton (57.365%)** and **VenusExpress (53.876%)** display a more balanced distribution, with ambiguity relatively evenly split between body and panel regions. Intermediate behaviors are observed for **Smart1** and **Soho**, where panel ambiguity remains dominant but less extreme.

Overall, these results indicate that ambiguity patterns are not uniform across satellite platforms and are strongly influenced by class-specific structural characteristics. This study was carried out to explain why panel was preferred over body in the hard assignment during inference.

4.4. Classification Performance and Confusion Matrix

The classification robustness is further confirmed by the confusion matrix in Fig. 5.

The model achieves a perfect classification of **2000/2000** instances for almost all classes, including *Cheops*, *LisaPathfinder*, *ObservationSat1*, *Proba2*, *Proba3*, *Proba3ocs*, and *Soho*. The **N/A** column (No Detection Occurred) reveals a minimal miss rate:

- **Smart1**: **1998** correct, **1** misclassified as *XMM Newton*, and **1** missed (*N/A*).
- **VenusExpress**: **1999** correct and **1** missed (*N/A*).
- **XMM Newton**: **1997** correct, **1** misclassified as *Smart1*, **1** as *VenusExpress*, and **1** missed (*N/A*).

The absence of values in the **N/A row** indicates **zero false positives**, confirming that the model never detects a satellite when none is present.

4.5. Test Set Results and Geometric Refinement

The final evaluation on the test set (Table 3) demonstrates the impact of the geometric refinement strategy.

Table 3. YOOLO metrics on the test set (Best Model, Conf 0.1).

Metric	Test Score
Classification (S_{cls})	0.9996
Detection (S_{det})	0.9703
Segmentation (S_{seg})	0.8748
Accuracy Score (S_{acc})	**0.938**
Final Score (S_{final})	**0.8564**

By deriving bounding boxes from the **minimal box circumscribing the segmented masks**, the detection score S_{det} reached **0.9703**. This, combined with a segmentation score S_{seg} of **0.8748** and a classification S_{cls} of **0.9996**, resulted in an accuracy score S_{acc} of **0.938** and a final score of **0.8564**, taking into account FLOPs and weights. Visual results on SPARK 2026 Challenge test set are shown in Fig. 6.

5. Conclusions

This work is based entirely on the SPARK 2026 dataset, specifically released for the current challenge. Direct comparisons with prior methods are inherently limited, as ex-

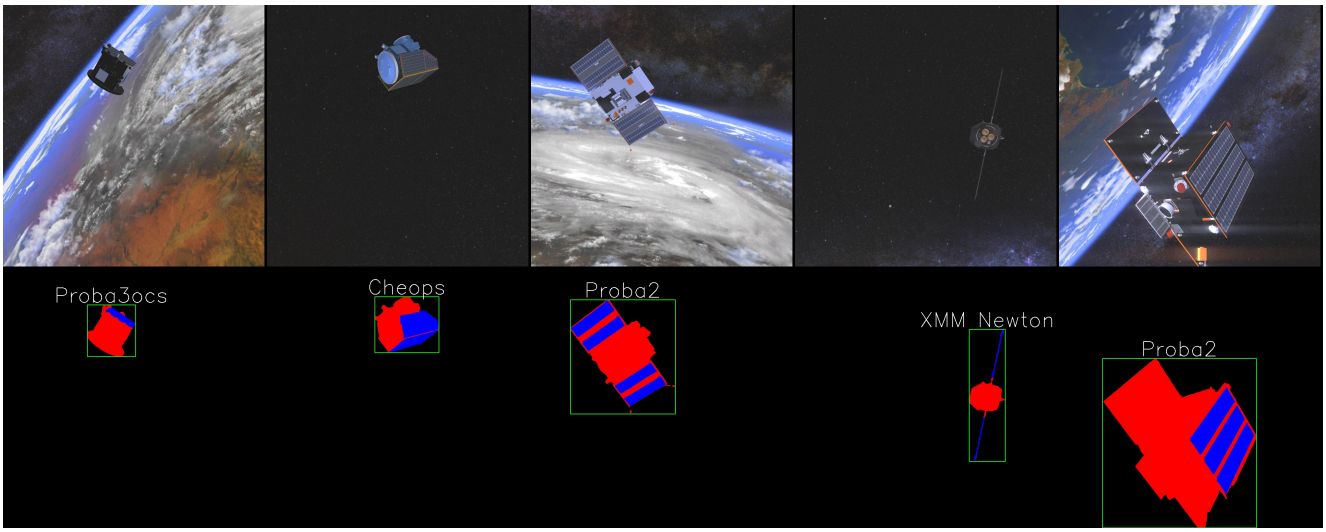

Figure 6. Visual results on SPARK 2026 Challenge test set for 5 spacecraft classes. In the first row the input images; in the second row the output images: in green the predicted bounding box, in red the predicted body masks, in cyan/blue the predicted panel mask (class names are written in white).

isting approaches were not evaluated under the same task formulation and protocol. For context, results were compared with those reported on previous SPARK dataset. Past research in space-object recognition has explored adapting one-stage detectors to the orbital domain. Notably, Zhou [17], using the SPARK 2022 dataset, built upon the YOLO11 architecture to improve bounding-box detection of orbiting spacecraft in synthetic imagery. Zhou's work focuses exclusively on detection, enhancing mAP through preprocessing techniques, loss-guided augmentation, and architectural modifications. In contrast, the SPARK 2026 challenge extends the task beyond detection, requiring simultaneous solutions for:

I. Image-level spacecraft classification,
II. Bounding-box detection,
III. Pixel-level semantic segmentation.

This extension shifts the focus from object localization to holistic scene understanding. While detection-only methods predict coarse object extents, segmentation provides dense spatial supervision, enabling learning of fine-grained geometric and structural features. The proposed YOOLO framework leverages this richer supervision to solve all three tasks within a unified YOLO11-Seg-based architecture. Despite the multi-task setting, detection performance remains highly competitive: Zhou [17] reported $mAP50-95$ = 0.926 and $mAP50$ = 0.99 using detection-only supervision, whereas YOOLO achieved $mAP50-95$ = 0.6778 and $mAP50$ = 0.8874 on SPARK 2026 validation. Importantly, YOOLO incorporates mask-level supervision for two spacecraft components (main body and panel), yielding strong segmentation performance: $mAP50-95$ = 0.8439 and $mAP50$ = 0.9867 on validation, and S_{seg} = 0.8505 on the test set. This demonstrates enhanced boundary consistency and structural awareness. Segmentation acts as an inductive bias, improving spatial coherence and alignment with true object boundaries. The YOLO11-Seg architecture integrates a shared backbone, a detection head for classification and bounding boxes, and a segmentation head for mask predictions. Training jointly on classification, detection, and segmentation losses allows YOOLO to maintain competitive detection accuracy while extending functional capability to segmentation. Despite the added segmentation branch, the model remains efficient, meeting the SPARK challenge constraints on parameters and computational cost. Overall, compared to detection-only solutions, YOOLO achieves a favorable trade-off between expressiveness and efficiency, providing a unified multi-task approach for spacecraft image understanding.

Acknowledgments

NANO BANANA 2, an AI-based tool, was used to assist in generating a schematic illustration of the proposed model architecture (Figure 1). The final figure was modified, reviewed and validated by the authors.

References

[1] N. Bobrinsky and L. Monte. The space situational awareness program of the european space agency. *Cosmic Research*, 48: 392–398, 2010. 1

[2] European Space Agency Space Debris Office. Esa's annual space environment report 2025. Technical report, European

Space Agency (ESA), 2025. ESA Space Environment Report 2025. Based on data collected through the end of 2024. 1

[3] Zijian He, Kang Wang, Tian Fang, Lei Su, Rui Chen, and Xihong Fei. Comprehensive performance evaluation of yolov11, yolov10, yolov9, yolov8 and yolov5 on object detection of power equipment. In *2025 37th Chinese Control and Decision Conference (CCDC)*, pages 1281–1286, 2025. 2

[4] Mate Kisantal, Sumant Sharma, Tae Park, Dario Izzo, Marcus Märtens, and Simone D'Amico. Satellite pose estimation challenge: Dataset, competition design and results, 2019. 1

[5] H. Klinkrad. *Space Debris: Models and Risk Analysis*. 2006. 1

[6] Yonghao Li, Hang Yang, Bo Lü, and Xiaotian Wu. Sld-yolo: A lightweight satellite component detection algorithm based on multi-scale feature fusion and attention mechanism. *Remote Sensing*, 17(17), 2025. 2

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *Computer Vision – ECCV 2014*, pages 740–755, Cham, 2014. Springer International Publishing. 3

[8] J.-C. Liou. An active debris removal parametric study for leo environment remediation. *Advances in Space Research*, 47 (11):1865–1876, 2011. 1

[9] Trupti Mahendrakar, Ryan T. White, and Madhur Tiwari. Spy: A context-based approach to spacecraft component detection, 2024. 2

[10] Federica Massimi, Pasquale Ferrara, and Francesco Benedetto. Deep learning methods for space situational awareness in mega-constellations satellite-based internet of things networks. *Sensors*, 23(1), 2023. 1

[11] Federica Massimi, Pasquale Ferrara, Roberto Petrucci, and Francesco Benedetto. Deep learning-based space debris detection for space situational awareness: A feasibility study applied to the radar processing. *IET Radar, Sonar & Navigation*, 18(4):635–648, 2024. 2

[12] Leo Pauly and Jamrozik. Lessons from a space lab: An image acquisition perspective. *International Journal of Aerospace Engineering*, 2023(1):9944614, 2023. 2

[13] Arunkumar Rathinam, Mohamed Adel Mohamed Ali, Vincent Gaudilliere, and Djamila Aouada. SPARK 2024: Datasets for Spacecraft Semantic Segmentation and Spacecraft Trajectory Estimation, 2024. 2

[14] Sumant Sharma and Simone D'Amico. Neural network-based pose estimation for noncooperative spacecraft rendezvous. *IEEE Transactions on Aerospace and Electronic Systems*, pages 1–1, 2020. 1

[15] Jianing Song, Duarte Rondao, and Nabil Aouf. Deep learning-based spacecraft relative navigation methods: A survey. *Acta Astronautica*, 191:22–40, 2022. 2

[16] Haiying Zhang, Zhengyang Li, and Chunyan Wang. Yolo-dynamic: A detection algorithm for spaceborne dynamic objects. *Sensors*, 24(23), 2024. 2

[17] Yi Zhou, Tianhao Zhang, Zijing Li, and Jianbin Qiu. Improved space object detection based on yolo11. *Aerospace*, 12(7), 2025. 7