Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI4Tech: AI Enabling Critical Technologies

SpaceDet: A Large-scale Space-based Image Dataset and RSO Detection for Space Situational Awareness

Jiaping Xiao[1†], **Rangya Zhang**[1†], **Yuhang Zhang**[1], **Lu Bai**[1*], **Qianlei Jia**[1] and **Mir Feroskhan**[1*]
[1]School of Mechanical and Aerospace Engineering, Nanyang Technological University
{jiaping001, rangya002, yuhang004}@e.ntu.edu.sg, jiaql@mail.nwpu.edu.cn, {bailu, mir.feroskhan}@ntu.edu.sg

Abstract

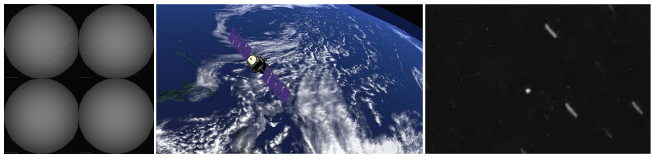
Space situational awareness (SSA) plays an imperative role in maintaining safe space operations, especially given the increasingly congested space traffic around the Earth. Space-based SSA offers a flexible and lightweight solution compared to traditional ground-based SSA. With advanced machine learning approaches, space-based SSA can extract features from high-resolution images in space to detect and track resident space objects (RSOs). However, existing spacecraft image datasets, such as SPARK, fall short of providing realistic camera observations, rendering the derived algorithms unsuitable for real SSA systems. In this work, we introduce SpaceDet, a large-scale realistic space-based image dataset for SSA. We consider accurate space orbit dynamics and a physical camera model with various noise distributions, generating images at the photon level. To extend the available observation window, four overlapping cameras are simulated with a fixed rotation angle. SpaceDet includes images of RSOs observed from $19km$ to $63,000km$, captured by a tracker operating in LEO, MEO, and GEO orbits over a period of $5,000$ seconds. Each image has a resolution of 4418×4418 pixels, providing detailed features for developing advanced SSA approaches. We split the dataset into three subsets: SpaceDet-100, SpaceDet-5000, and SpaceDet-full, catering to various image processing applications. The SpaceDet-full corpus includes a comprehensive dataloader with $781.5GB$ of images and $25.9MB$ of ground truth labels. Furthermore, we adapted detection and tracking algorithms on the collected dataset using a specified splitting method to accelerate the training process. The trained model can detect RSOs from real-world space observations with zero-shot capability.

1 Introduction

Space Situational Awareness (SSA) [Wang *et al.*, 2022] plays a crucial role in ensuring the safety of space assets by providing real-time information perception and risk evaluation for space operations, such as spacecraft navigation [Hein, 2020] and debris mitigation [Usovik, 2023]. Conventional SSA systems, like those used by the Japanese Space Agency (JAXA) [Harris *et al.*, 2021], rely on observing resident space objects (RSOs) and determining their orbits using ground-based facilities equipped with large telescopes and radars. These systems necessitate extensive site areas, high costs, and specific geographical locations [Jia *et al.*, 2024]. Given the computational limitations of satellites, current space-based SSA systems, which involve complex numerical calculations, typically depend on the space-ground network for data processing and information fusion. This reliance results in substantial communication loads and delays.

With the advancements in artificial intelligence (AI) [Pisutsin *et al.*, 2024] and high-performance edge computing [Xiao *et al.*, 2024], an onboard vision-based SSA system presents a more flexible and lightweight alternative to traditional ground-based SSA for RSO detection and tracking. One of the primary challenges in SSA is providing accurate position and orientation vectors (observations) of targets to determine their orbits. Methods such as Gauss's method [Vallado, 2001], which requires at least three observations for preliminary orbit determination, and Lambert's method [Engels and Junkins, 1981], which needs only two position vectors with temporal information, are used for this purpose. Essentially, increasing the number of observations enhances the accuracy of orbit determination, highlighting the importance of the object detection and tracking (ODT) component in SSA. To develop precise and practical ODT algorithms, extensive high-resolution space imagery is essential. However, most existing spacecraft image generation techniques [Musallam *et al.*, 2021a] rely on high-fidelity simulators that ignore space camera models and the cosmic background, resulting in unrealistic images (see Figure 1) and ODT algorithms unsuitable for real SSA systems.

In this work, we present SpaceDet, a large-scale realistic space-based image dataset for SSA (project in https://github.com/NTU-ICG/SpaceDet). This dataset considers accurate space orbit dynamics and a physical camera model with various noise distributions, generating images at the photon level. To extend the observation window, we simulate four overlapping cameras with a fixed rotation angle. SpaceDet comprises images of RSOs observed from distances ranging from 19 km to $63,000$ km, captured by a tracker operating in Low Earth Orbit (LEO), Medium Earth Orbit (MEO), and Geostation-



(a) (b) (c)

Figure 1: Comparison of our SpaceDet images with SPARK images [Musallam *et al.*, 2021a] and real-life observed images. (a) SpaceDet images at timestamp 0 (four cameras from left top to right bottom), which show the realistic exposure with noise distribution; (b) A simulated spacecraft image from SPARK; (c) The real-life space observation image from the telescope and sensor network (EGTN[1]). The similar streaks due to the exposure of fast-moving RSOs and the hot pixels induced by the noises in (a) and (c) demonstrate the realistic images in our SpaceDet dataset.

ary Orbit (GEO) over $5,000$ seconds. Each image boasts a resolution of 4418×4418 pixels, providing detailed features for the development of advanced SSA approaches. We have divided the dataset into three subsets: SpaceDet-100, SpaceDet-5000, and SpaceDet-full, catering to various image processing applications. The SpaceDet-full corpus includes a comprehensive dataloader with 781.5 GB of images and 25.9 MB of ground truth labels. To the best of our knowledge, SpaceDet is the first image dataset to offer four-camera observations with realistic image generation from space for space ODT. The key contributions and features of this dataset are summarized as follows:

Realistic Image Generation: Incorporating accurate space orbit dynamics and a physical camera model with various noise distributions to produce photon-level realistic space images.

Multiple Camera Observations: Simulating four overlapping cameras with fixed rotation angles to extend the observation window.

Large Range Tracker Observation: Covering RSO images observed from 19 km to $63,000$ km for the tracker operating in LEO, MEO, and GEO orbits.

Automated Label Generation with Bearing Angle: Providing accurate ground truth labels with bearing angle information generated by the simulator through an automated transformation and annotation process.

Extensive Benchmarks: Benchmarking the dataset using state-of-the-art (SOTA) algorithms, including YOLOv5, YOLOv8, YOLOv10, DINO, etc., on SpaceDet-100 with a specified splitting method to expedite the training process. Additionally, various object tracking methods are compared on SpaceDet-100 to explore its applications.

2 Related Dataset Work

Publicly available image datasets for space object imagery are predominantly ground-based, such as SatNet and SatSim [Fletcher *et al.*, 2019]. Existing space-based image datasets, such as BUAA-SID-POSE 1.0 [Qiao *et al.*, 2022], SPEED [Kisantal *et al.*, 2020], SPEED+ [Park *et al.*, 2023], and URSO [Proença and Gao, 2020], primarily emphasize spacecraft pose estimation [Pauly *et al.*, 2023]. These datasets typically feature a limited number of RSOs in the images and lack comprehensive annotations such as bounding boxes, which are essential for broader SSA applications beyond pose estimation. Since space-borne real data is often challenging and expensive to acquire, simulated datasets have become the predominant approach for developing methods for SSA tasks. BUAA-SID-share 1.0 [Zhang *et al.*, 2010] features various satellite models created using 3dsMax but lacks simulation of the space environment. The SPARK [Musallam *et al.*, 2021a] dataset includes simulated models of different satellites and space debris but lacks realistic camera observations. An annotated dataset derived from the RSO Near-Space Astrometric Research (RSONAR) mission is provided by [Suthakar *et al.*, 2023], which collected data using a low-resolution, wide-field-of-view imager on a stratospheric balloon.

Dataset	#Images	#Objects	Resolution	Object/Image	Public?
BUAA-SID-share 1.0	9.2k	20	320×240	single	yes
SPARK	30k	11	1440× 1080	single	request
RSONAR	429	3	1024× 1024	multiple	no
SpaceDet-100	100	56	4418×4418	multiple	yes
SpaceDet-5000	5k	414	4418×4418	multiple	yes
SpaceDet-full	20k	673	4418×4418	multiple	yes

Table 1: Comparisons of SpaceDet with existing datasets.

Additionally, some datasets have been generated by researchers to simulate space conditions and RSOs, facilitating algorithm development and testing [Tang *et al.*, 2023; Chen *et al.*, 2023; Shen *et al.*, 2024]. However, these datasets are often inaccessible and lack comprehensive reality analysis. Table 1 provides a summary of statistics for existing space-based RSO detection image datasets as well as our SpaceDet dataset. SpaceDet captures more RSOs in the images and has a higher resolution compared to prior datasets.

3 Data Curation Process

3.1 Data Generation

The SpaceDet dataset is collected from a real-time high-fidelity simulator based on precise space orbit dynamics and physical camera models. Since the space-based observer operates at an altitude of 500 km, the effects of the atmosphere and related noise are not included in the modeling process. The space environment model simulates a catalog of RSOs in orbit around the Earth, along with other celestial bodies in the sky. The RSO simulation is based on the United States 18th SDS Space Catalog [2], which is fetched in Two-Line Elements (TLEs) format for the desired simulation epoch and propagated using an SGP4 propagator. The propagator provides the positions and velocity vectors of all objects in the TEME coordinate system, which is used to populate the 3D environment.

The modeling of environmental noise expected for a sensor is also incorporated. Under favorable imaging conditions, the sensor's payload is oriented away from the sun and perpendicular to the orbit, allowing the primary background noise

[1]https://exoanalytic.com/space-domain-awareness

[2]http://space-track.org/

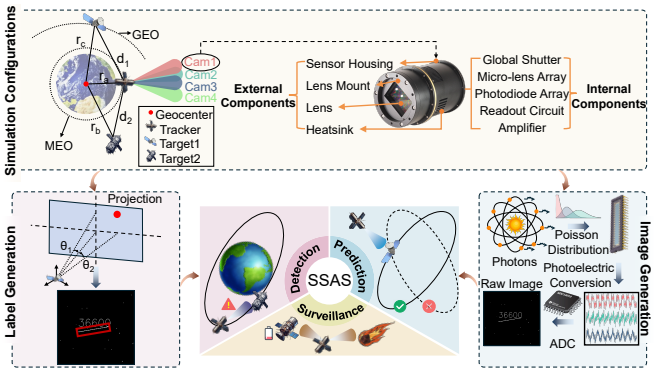


Figure 2: Overall framework of the data curation process.

source to be the Earth's limb—the bright edge of the Earth's horizon. This background noise is modeled using data from the Hubble Space Telescope [3] and the NEOSSAT mission [Thorsteinson, 2018]. The camera captures a circular image on the image plane, and the detector on the focal plane records the digital image.

To detect which RSOs crossing the field of view (FOV) can be identified by the sensor, a photometric detection model is applied. The sensor detects an object only when its signal-to-noise ratio (SNR) exceeds a specified threshold, typically set at 5. The received signal is calculated assuming 100% diffuse reflection, where the fraction of incident sunlight reflected to the sensor is given by:

$$\text{Reflection Factor} = \frac{2\mu r^2}{3\pi R^2} \times (\sin\phi + (\pi - \phi)\cos\phi), \quad (1)$$

where μ is the object's reflectivity, r is the object's radius, R is the distance between the object and the telescope, and ϕ is the phase angle of reflection. The solar flux, considered as the solar constant at 1 AU from the sun, is assumed to be uniform across all objects, as the variation in distance from the sun is negligible. The brightness data from the star catalog is used to determine the signal level from stars, which is distributed across a Gaussian spot formed on the detector.

The model for projecting star/object positions onto the image is based on a pinhole camera model, supplemented by a compound distortion model (radial and tangential), similar to the Brown-Conrady model [Zhang, 2000]. To calculate the noise electrons, three noise sources are modeled: shot noise with a random distribution, sensor noise (e.g., dark current) modeled with a Poisson distribution, and read noise modeled with a normal distribution. Additionally, background noise is modeled with a Poisson distribution. Relative illumination is implemented as a quadratic function to account for roll-off and vignetting effects in the image. These signal and noise models provide the number of photoelectrons collected by each pixel on the sensor, which are then converted into 16-bit digital values (0-65535). Note that four cameras (60°, 75°, 90° and 105° azimuth angle for Cam1 to Cam4, respectively) are adopted to generate the images simultaneously.

[3]https://hst-docs.stsci.edu/stisihb/chapter-6-exposure-time-calculations/6-5-detector-and-sky-backgrounds

The overall framework for the data curation process is illustrated in Figure 2. The datasets generated by the simulator are in the forms of images (*TIFF format*) and a set of metadata (*CSV format*). All *state information* (including position, velocity and attitude) of RSO is propagated with the public in-catalog TLEs. The data collection is free of any ethical issue or participation risk. The space orbit propagation program is developed based on the Standards of Fundamental Astronomy (SOFA) [IAU, 2021] package and SGP4 model [Vallado and Crawford, 2008]. With camera specifications such as lens parameters, sensor parameters, and camera pointing direction angles (elevation and azimuth angles in the RSW coordinate frame), the physical camera model can generate *pixel values* of images at each timestep based on the aforementioned relative illumination and noise distributions. The physical camera model used for generating realistic space-based images includes several key components, including the pinhole camera model, lens distortion, and noise modeling. Each 3D point $\mathbf{X}$ in the space is projected onto the image plane using the pinhole camera model, then distorted based on the lens distortion model, and finally, various noise distributions as mentioned are added to simulate the physical conditions of space imaging. For instance, we model the noise as a combination of Poisson noise (sensor noise) and Gaussian noise (read noise):

$$I_{noisy}(u,v) = \text{Poisson}(I(u,v)) + \mathcal{N}(0,\sigma^2), \quad (2)$$

where $I(u,v)$ is the intensity value at pixel (u,v), $\text{Poisson}(I(u,v))$ represents the Poisson noise, and $\mathcal{N}(0,\sigma^2)$ represents the Gaussian noise with mean 0 and variance σ^2. The exposure time (*1 second*) is reflected in the image generation as we overlap the images over the exposure time into one image. The ground truth bounding box is calculated with the bearing angles (θ_1 and θ_2) of a target with respect to the tracker as defined in Figure 2. By selecting the starting time (*YYYY-MM-DD HH:MM:SS* in UTC) and simulation duration, we can generate desired images and metadata over a certain period. The specified simulation time is from *2023-01-01 0:00:00* to *2023-01-01 1:23:20* with a 1-second time difference for the successive images (this time difference is the exposure time and is optimized for object detection).

3.2 Dataset Validity and Uniqueness

Currently, there are fewer than six datasets available in this field, and they are all based on simulations, as NASA's database is not publicly accessible. The dataset presented in this work is the first large-scale, realistic, space-based image dataset at the photon level, aiming to bridge the gap between simulated and real-world data. Most existing datasets, such as BUAA-SID-share 1.0 [Zhang *et al.*, 2010], SPARK [Musallam *et al.*, 2021b], and the Space Target Dataset [Zhang *et al.*, 2022c], are primarily generated for satellite pose estimation and space target classification in ideal simulation conditions. These datasets focus on capturing targets from close distances and multiple angles to emphasize single-target characteristics. In contrast, our dataset captures targets at various distances based on realistic space-based camera observations. We have compared our images with the real observed images with a ground-based telescope (see project website), where the starfield and captured positions at various timestamps are

first compared to show the accuracy of the simulator. These imaging results then clearly validate the realistic characteristics of our simulated images.

3.3 Data Annotation

All images in the SpaceDet dataset are annotated with classes indicating LEO, MEO, and GEO (low, medium, and high accordingly), as well as 2D bounding boxes for the labeled parts (see Figure 3). To ensure high-quality annotations, all classes and bounding box information are automatically derived from the orbital and positional information of space objects, rather than being manually labeled. All the orbital and positional information of these space objects is generated from the aforementioned orbit propagation simulator containing a semi-major axis (SMA) and two bearing angles.

Classes Annotation: SMA, a key parameter for describing orbital ellipses, determines the size and shape of the orbit. Targets are classified as LEO (SMA $\leq$ $8413km$), MEO ($8413km$ $<$ SMA $\leq$ $42240km$) , and GEO (SMA $>$ $42240km$) based on their SMA.

Bounding Box Annotation: Bounding boxes are derived from two bearing angles θ_1 θ_2 of space objects. The bearing angle information is defined in the camera frame, while the pixel position is defined in the pixel coordinate system (origin at the upper left, x-axis to the right, y-axis downward). The transformation from bearing angles to pixel positions is given as follows:

$$x_{\text{pixel}} = \left(\tan(\theta_1) \cdot \frac{\text{focal\_length}}{\text{tb\_number}} + 0.5 \right) \cdot \text{width}, \quad (3)$$

$$y_{\text{pixel}} = \left(\tan(\theta_2) \cdot \frac{\text{focal\_length}}{\text{tb\_number}} - 0.5 \right) \cdot (-\text{width}), \quad (4)$$

where tb_number is the effective sensor size, and width is the number of pixels in the image's width. The transformations provide the position of the space object at a specific moment in the image. Since each image has a 1-second exposure time, the objects appear as a path showing their movement during that second. To find the bounding box, we calculate the pixel coordinates of the space object at the start and end of the exposure. These two points form the diagonal corners of the bounding box, which helps us accurately determine the size and location of each bounding box.

3.4 Image Slicing

The original image size is 4418×4418 far beyond the processing capabilities of YOLO and most GPUs. To address the issue of large image dimensions, the images are initially sliced into smaller sections measuring 260×260 pixels each (see Figure 3 (c)). Multiple split sizes were tested, and 260x260 was selected as the optimal size based on experimental results. This slicing process includes an adjustable overlap in both horizontal and vertical directions, serving as a strategy for data augmentation. The last slices are aligned to the image's edge and then cut to the predefined size. Additionally, the annotations of the labels are accurately adjusted to match the newly sliced dimensions. Given the inherent

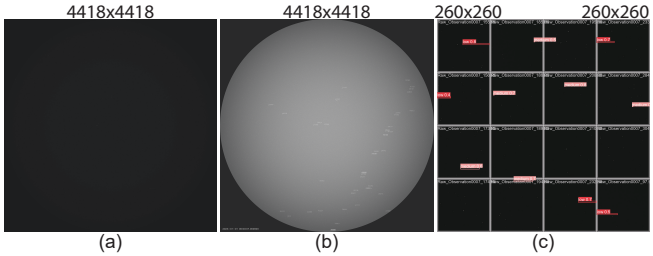


Figure 3: The image sample from the SpaceDet dataset. (a) The original image generated from the simulator with 1 second exposure time; (b) The post-processed image to show the observed image and RSOs with reference IDs; (c) The sliced image batch and annotations.

sparse labeling of the dataset, 96% of the sliced images do not contain the target. Addressing the risk of overfitting caused by a high proportion of negative samples (images without targets), the training and validation datasets are selectively pruned to remove a substantial number of negative samples. This strategy is directed towards creating a more balanced dataset, aiming for an approximate 0.9 : 0.1 ratio between images with and without targets. For the test dataset, selective pruning is omitted to maintain the accuracy and validity of model evaluation.

3.5 Dataset Release

The release of the SpaceDet dataset is structured into three distinct subsets, namely SpaceDet-100, SpaceDet-5000, and SpaceDet-full (see Table 1), to cater to varying levels of image processing and analysis requirements.

SpaceDet-100: This is the minimal dataset intended for preliminary training and testing purposes. It includes 100 high-resolution images that provide a foundational basis for algorithm development and initial performance assessments. This subset is ideal for quick iteration cycles and for researchers who are beginning their work on SSA without requiring extensive computational resources. SpaceDet-100 is particularly useful for initial model training and validation, performance benchmarking of new methods, and educational purposes, allowing students and new researchers to get hands-on experience with SSA data.

SpaceDet-5000: This subset expands the dataset to 5000 images, all captured from Camera 2. It is designed to offer a more comprehensive dataset that can be used for more rigorous training and testing of machine learning models. SpaceDet-5000 provides a larger sample size to improve the robustness of algorithms and to ensure that the models are exposed to a wider variety of scenarios and conditions encountered in space-based observations. It is intended for detailed algorithm development and refinement, robustness testing across a larger set of scenarios, and intermediate-scale projects that require significant but manageable computational resources.

SpaceDet-full: This is the full version of the SpaceDet dataset, featuring 5000 images captured from each of the four simulated cameras, resulting in a total of 20,000 images. This

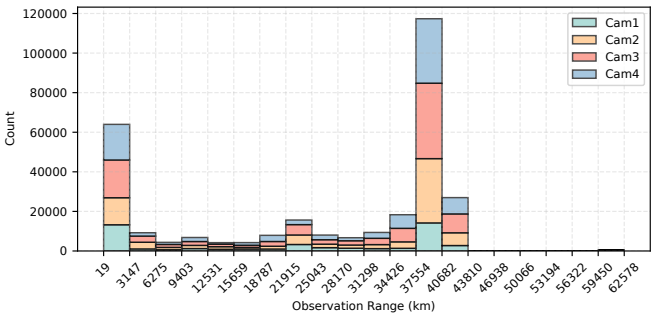


Figure 4: Histogram of observation range for four cameras in SpaceDet-full.

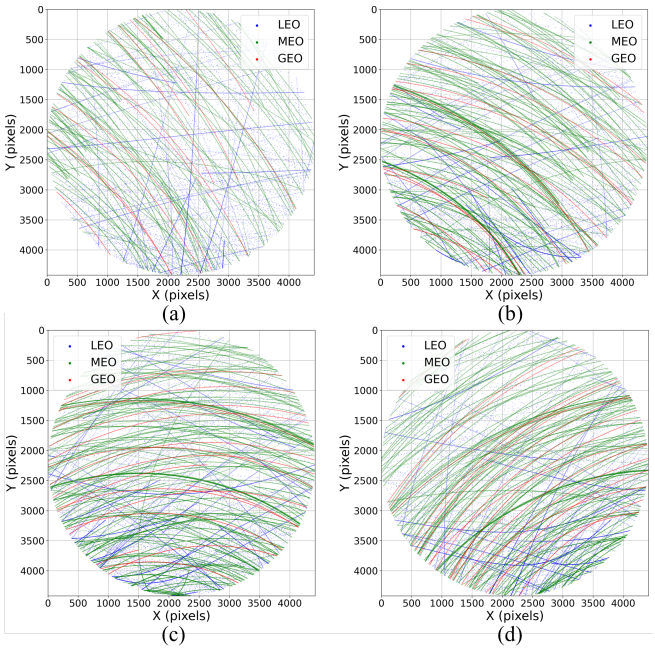


Figure 5: Positional distribution of targets across four camera datasets, categorized by orbital type (LEO, MEO, GEO). (a) Camera 1; (b) Camera 2; (c) Camera 3; (d) Camera 4.

comprehensive dataset is collected for advanced research and development objectives. It supports the training and validation of complex models that require multi-view observations to accurately detect and track RSOs. The multi-camera setup allows researchers to develop and test algorithms capable of leveraging spatial information from different perspectives, enhancing the accuracy and reliability of SSA systems.

4 SpaceDet Dataset Analysis

Statistical Features of Individual Images: Here, we provide a specific quantitative description of images of size 4418×4418 pixels (*16-bit*, *39.1MB* in storage). After testing, the average signal-to-noise ratio (SNR) and average root mean square (RMS) contrast [Peli, 1990] of SpaceDet-full images are 1.94 dB and 4.67, respectively (typical values for general images are 30 dB for SNR and 80 for contrast). With a threshold pixel value of 2000, the bright point ratio is

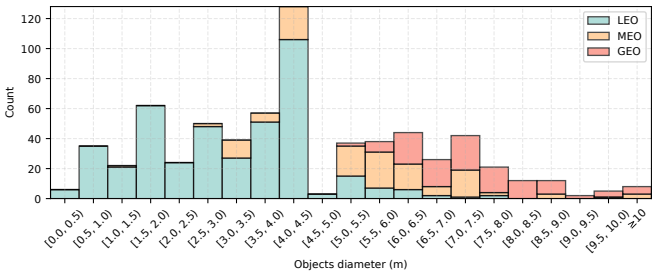


Figure 6: The size distributions for all observed objects categorized by orbital type (LEO, MEO, GEO)

0.47. This large image size ensures high resolution and clarity, which is beneficial for detailed analysis and visualization in various applications. However, hardware or applications such as edge computing in satellites may have difficulty handling images with such high resolution. In such cases, we suggest compressing the images before using this dataset. In real-space environments, various types of noise, such as optical and electromagnetic interference, can be present, which manifest in the SNR parameter of this dataset. Generally, a smaller SNR indicates a greater amount of noise in the images (see Figure 3 (b)). This suggests that the image information in this dataset more closely aligns with real-world conditions but poses challenges for feature extraction. Additionally, due to variations in the distance between the targets and the cameras, the brightness of different targets varies, which is a natural phenomenon in imaging. The targets appear brighter when they are closer to the camera and dimmer when they are farther away. As some applications may have specific requirements for image brightness, some image processing algorithms may be needed to enhance brightness.

Statistical Features of the Whole Dataset: Figure 4 shows the histogram of observation ranges for four cameras in SpaceDet-full. The horizontal axis of the figure represents the observation distance of the cameras, and the vertical axis represents the counts of target occurrences. Each bar in the histogram corresponds to a range of 3128 km. The closest observed target is at a distance of 19 km, while the farthest is at 62,578 km. The bars for the ranges 19 to 3147 km and 37,554 to 40,682 km are the highest, indicating the highest frequency of target occurrences. In contrast, the frequencies are much lower for other ranges, resembling a long-tail distribution commonly seen in the dataset.

Figure 5 shows the positional distribution of targets across four camera datasets. In these maps, the origin of the coordinates is set at the top-left corner of each image, with x and y representing pixel coordinates. The targets are distinguished by their orbital categories: blue points for LEO targets, green points for MEO targets, and red points for GEO targets. It can be observed that the red and green points form dense, curved trajectories. This is because MEO and GEO targets, being farther from the observer, appear to move more slowly in the images, resulting in closely spaced positions across consecutive frames. In contrast, LEO targets are more widely dispersed, indicating a faster apparent movement across the images. The figure also shows that there is no discernible pattern in the

occurrence and density of the targets. The complex nature of the space environment, combined with variations in observation times, viewing angles, and the orbital paths of tracking instruments, leads to differences in target position distributions. This complexity poses a significant challenge for space object detection.

Since the orientations of the objects are negligible considering the far-range observation, the size distribution for all observed objects is illustrated in Figure 6. From all distributions, most of the RSOs are within 0.5m to 10m, and the smallest object is approximately 30cm. Given the large detection range, detecting such small objects in space is extremely challenging, which highlights the significant value of our dataset and benchmark pipeline.

5 Benchmark

As SpaceDet is regarded as a new SSA dataset, we provide several ODT baseline results based on representative one or two stage detectors and detection-based multiple object tracking methods on SpaceDet-100. All the dataset and the code are available in https://github.com/NTU-ICG/SpaceDet. Additionally, detailed descriptions of the experimental settings and additional comparison results can be found in the supplementary material accordingly.

5.1 Basic Settings

We utilize the Ultralytics library [Jocher *et al.*, 2023] under the AGPL-3.0 license, which includes all YOLO series models for ODT. Other architectures are implemented using the MMDetection toolbox [Chen *et al.*, 2019] from OpenMMLab. Following standard configurations, all models are trained on an NVIDIA RTX A6000 GPU (48GB) for up to 400 epochs with an Intel Xeon w9-3495X CPU (4.8 GHz), with early stopping if there is no improvement for 30 epochs. From the SpaceDet-100 dataset, 70% of images are used for training and 20% for validation. To minimize randomness in testing, 100 images from the SpaceDet-5000 dataset are selected for evaluation. Metrics for detection performance include `mAP@50`, `mAP@50-95`, `precision`, `recall`, and `F1 score`, while object tracking is assessed using `false positives`, `ID switches`, and `multiple object tracking accuracy` (MOTA).

5.2 Object Detection Benchmark

In our object detection benchmark, we evaluated various models, primarily YOLOv8m. We also included yolov3_mobilenetv2, using MobileNetV2 as the YOLOv3 backbone [Redmon, 2018], and faster_rcnn, based on the ResNet architecture [Ren *et al.*, 2016]. Transformer-based models such as DETR [Carion *et al.*, 2020], Deformable-DETR [Zhu *et al.*, 2020], and DINO [Zhang *et al.*, 2022a] were also employed. To optimize training, batch sizes were adjusted based on GPU memory constraints. All models utilized our custom preprocessing pipeline, tailored to the challenges of SSA datasets. All models used 260x260 images as input to ensure result reliability.

The experimental results in Table 2 show the strengths and weaknesses of each method in terms of model size, speed, and accuracy. Overall, existing SOTA methods struggle in the space domain due to the sparse distribution of small targets, many of which occupy only a few pixels in large 4418×4418 images. Over 99% of each image is background noise, presenting challenges not found in typical detection tasks, which emphasizes the unique value of our dataset. Transformer-based models capture complex features but require more memory and inference time, making them less suitable for space-based tasks. Faster_rcnn achieves an acceptable level of accuracy but demands significant memory and long detection times, indicating a trade-off between accuracy and efficiency. Yolov3_mobilenetv2 provides faster detection but sacrifices performance, highlighting the challenges of balancing speed and accuracy in SSA. The YOLOv8m model stands out with its small size, high accuracy, and reasonable detection speed, making it the most suitable for SSA tasks. Its balance of speed and precision effectively meets the unique challenges of space-based monitoring.

Model Info	Training Process					Testing Process			
Models	batch	Mem	T/epoch	Epochs	Size(MB)	P	R	F1	T/img
yolov3_mobilenetv2	48	34.36G	5.09s	57	35.9	0.288 ± 0.005	0.277 ± 0.023	0.282 ± 0.010	**1.99s**
faster_rcnn	40	42.70G	1.82s	24	333.8	0.347 ± 0.013	0.315 ± 0.020	0.329 ± 0.009	6.42s
DETR	8	23.18G	1.78s	209	512.2	0.236 ± 0.055	0.312 ± 0.026	0.267 ± 0.043	4.60s
deformable_detr	8	38.91G	4.96s	141	498.8	0.315 ± 0.017	0.479 ± 0.024	0.380 ± 0.018	9.71s
DINO	8	34.13G	6.62s	35	597.7	0.332 ± 0.014	0.495 ± 0.092	0.394 ± 0.021	13.15s
YOLOv8m	48	19.20G	3.81s	209	52.1	0.600 ± 0.017	0.435 ± 0.005	**0.492 ± 0.019**	3.81s

Table 2: Performance comparison of SOTA models for space object detection (averaged over three runs). Mem denotes GPU memory usage during training. T/epoch refers to training time per epoch. Epoch indicates the number of epochs until convergence. Size refers to the storage size of trained models. Precision (P), recall (R), and F1 score are presented as mean ± standard deviation.

In selecting SOTA models for space object detection, the YOLO series was chosen for its strong performance across various tasks. Given the demands of space deployment, factors such as computational complexity, detection speed, and accuracy are crucial. Various YOLO versions and sizes offer trade-offs among these factors. To identify the optimal model for space deployment, comprehensive experiments were conducted. As shown in Table 3, three object detection models are compared: YOLOv5 [Jocher, 2020], YOLOv8 [Jocher *et al.*, 2023], and YOLOv10 [Wang *et al.*, 2024], each evaluated with five parameter sizes (n, s, m, l, x). Generally, larger models (m, l, x) achieve better accuracy but require more memory and longer training times. While YOLOv8m demonstrated superior performance, smaller models like YOLOv5n and YOLOv8s exhibited lower performance but faster detection times. YOLOv8n balanced speed and accuracy effectively, making it more suitable for SSA, where lightweight computation and timely detection are essential.

Our experiments on the SpaceDet-100 dataset reveal that existing SOTA ODT methods, effective in conventional scenarios, underperform in space environments, achieving significantly lower scores compared to standard datasets. This underscores the need for algorithms specifically designed for SSA, and our dataset aims to bridge this gap.

5.3 Object Tracking Benchmark

SpaceDet can also be used to evaluate multiple object tracking methods, as it contains ID information for each space object. The ground truth tracking data includes timestamps, ob-

Model Info	Training Process				Testing Process			
Models	Mem	T/epoch	Epochs	Size(MB)	P	R	F1	T/img
v5n	**6.38G**	**8s**	62	5.30	0.78	0.18	0.29	**1.93s**
v5s	10.20G	26s	312	18.60	0.66	0.33	0.44	2.36s
v5m	18.20G	25s	226	50.50	0.72	0.26	0.38	3.45s
v5l	28.00G	37s	236	106.80	0.65	0.33	0.44	4.88s
v5x	40.80G	63s	239	195.00	0.65	0.33	0.44	7.07s
v8n	6.65G	8s	236	6.30	0.65	0.36	**0.47**	2.01s
v8s	10.60G	13s	72	22.50	0.72	0.18	0.29	2.37s
v8m	19.20G	25s	228	52.10	0.62	0.38	**0.47**	3.81s
v8l	29.80G	41s	305	87.70	0.55	0.38	0.45	5.55s
v8x	37.40G	61s	192	136.70	0.58	0.38	0.46	7.02s
v10n	9.10G	11s	345	5.80	0.65	0.28	0.39	2.38s
v10s	15.60G	19s	85	16.50	0.67	0.27	0.38	2.63s
v10m	26.30G	31s	216	33.50	0.63	0.34	0.44	3.74s
v10l	40.80G	48s	188	52.20	0.66	0.34	0.45	5.66s
v10x	41.80G	68s	242	64.10	0.65	0.36	0.46	6.54s

Table 3: Detection results of baseline YOLO models on SpaceDet-100. During the training of v10x, in order to prevent the GPU memory from being full, we adjusted the batch size to 36.

ject IDs, and two bearing angles. During tracking, pixel positions of the targets are converted into bearing angles using Eq. (3) and (4) for comparison with the actual data.

Table 4 summarizes the results of various tracking methods based on a YOLOv8n model with an F1 score of 0.4674. Two tracking methods were tested: ByteTrack [Zhang *et al.*, 2022b] and BoT-SORT [Aharon *et al.*, 2022], using Intersection over Union (IoU) and Euclidean distance for similarity calculations. Variants of BoT-SORT incorporated different global motion compensation algorithms (ECC, ORB, SIFT, and Sparse Optical Flow), and feature-based similarity calculations were explored using features from the pre-trained YOLOv8n model, as well as traditional methods like HOG and SIFT. From Table 4, it is evident that the performance of models using IoU distance is far inferior to those using Euclidean distance. This is because space targets are small and fast-moving, and even slight calculation errors can result in an IoU of 0. For fast-moving objects in space, minor camera movements have minimal impact on tracking. Consequently, the evaluation metrics for different global motion compensation methods show little variation in MOTA. BoT-SORT, an improved version of ByteTrack with global motion compensation, performs similarly to ByteTrack because the SSA dataset is insensitive to camera motion. For the SpaceDet dataset, global motion compensation increases processing time with minimal benefits. Although IoU and Euclidean distance calculations are fast, their accuracy is lower than feature distance based on YOLO. Among all tracking methods, the YOLO feature extraction method performs the best, with a 27% higher accuracy than Euclidean distance. Additionally, traditional feature extraction methods like HOG and SIFT perform poorly.

In addition to the limitations on computing resources and speed, space target tracking tasks must also minimize false detections and ID switches. Excessive false detections and ID switches can negatively impact subsequent tasks such as orbit determination and orbit propagation. BoT-SORT, based on YOLO feature extraction, performs well in all these aspects, making it more suitable for SSA tasks.

Model Info	Target Number		Object Number		Evaluation Metrics					
Models	Total	Predict	Real	Predict	Matches	Misses	FP	IDs	MOTA	Time
Byte_iou	4695	182	56	180	178	4517	4	143	0.0065 ± 0.0002	0.06
Byte_euclidean	4695	2111	56	134	2098	2597	13	202	0.4012 ± 0.0072	0.05
BoT_iou	4695	182	56	181	178	4517	4	143	0.0065 ± 0.0002	0.05
BoT_euclidean	4695	2107	56	150	2095	2600	11	183	0.4049 ± 0.0078	0.05
BoT_euclidean_ecc	4695	2101	56	145	2092	2603	9	203	0.4002 ± 0.0075	1.53
BoT_euclidean_orb	4695	2105	56	145	2095	2597	9	204	0.4003 ± 0.0082	0.10
BoT_euclidean_sift	4695	2102	56	139	2092	2603	10	196	0.4012 ± 0.0085	1.15
BoT_euclidean_sparse	4695	2106	56	145	2096	2599	10	202	0.4017 ± 0.0089	0.14
BoT_feature_yolo	4695	2498	56	53	2486	2209	12	51	**0.5160 ± 0.0091**	0.26
BoT_feature_hog	4695	499	56	57	495	4200	3	52	0.0938 ± 0.0044	7.46
BoT_feature_sift	4695	121	56	33	117	4578	3	4	0.0235 ± 0.0010	1.42

Table 4: Performance evaluation of multiple object tracking methods on SpaceDet-100 (averaged over three runs). Total and Predict columns refer to the total and predicted target numbers, respectively. Evaluation metrics include Matches, Misses, False Positives (FP), ID switches (IDs), MOTA, and tracking time (Time).

6 Limitations of SpaceDet Dataset

Despite the comprehensive nature of the SpaceDet dataset, there are several limitations to consider. First, while the dataset is designed to be highly realistic, the images are still generated via simulations, which may not capture all the complexities and variabilities of real-space environments. Second, the dataset focuses on high-resolution images, which, while beneficial for detailed analysis, also require considerable computational resources for processing and storage, potentially limiting accessibility for researchers with limited resources. Lastly, while the dataset includes a range of orbital distances and conditions, it hasn't covered all possible scenarios that SSA systems might encounter, necessitating further validation with real-world data to ensure robustness and generalizability of the developed algorithms.

7 Conclusion

Focusing on improving SSA, we present SpaceDet, a large-scale realistic space-based image dataset designed to overcome the limitations of existing datasets such as SPARK. SpaceDet provides a comprehensive collection of high-resolution images (4418×4418 pixels) generated using accurate space orbit dynamics and a physical camera model with Poisson noise distribution, capturing observations from 19 km to 63,000 km. The dataset is divided into three subsets: SpaceDet-100, SpaceDet-5000, and SpaceDet-full, each catering to different research needs. Our benchmark evaluations show that while larger YOLO models generally outperform smaller ones, lightweight models like YOLOv5s and YOLOv8n offer faster detection speeds, crucial for space-based applications with limited computing resources. Moreover, SOTA ODT methods perform inadequately in the space environment, highlighting the necessity for algorithms tailored to SSA. SpaceDet not only facilitates the development of new ODT algorithms but also serves as a benchmark for evaluating advanced SSA techniques.

Acknowledgments

This work was supported by the Economic Development Board (EDB) Singapore under its Space Technology Development Programme (STDP) Thematic Grant Call on Space Technologies (Award S23-020019-STDP).

Contribution Statement

†Jiaping Xiao and Rangya Zhang are co-first authors; *Lu Bai and Mir Feroskhan are corresponding authors.

References

[Aharon *et al.*, 2022] Nir Aharon, Roy Orfaig, and Ben-Zion Bobrovsky. Bot-sort: Robust associations multi-pedestrian tracking. *arXiv preprint arXiv:2206.14651*, 2022.

[Carion *et al.*, 2020] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *European conference on computer vision*, pages 213–229. Springer, 2020.

[Chen *et al.*, 2019] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.

[Chen *et al.*, 2023] Zhe Chen, Yang Yang, Anne Bettens, Youngho Eun, and Xiaofeng Wu. A simulation-augmented benchmarking framework for automatic rso streak detection in single-frame space images. *arXiv preprint arXiv:2305.00412*, 2023.

[Engels and Junkins, 1981] RC Engels and John L Junkins. The gravity-perturbed lambert problem: A ks variation of parameters approach. *Celestial mechanics*, 24(1):3–21, 1981.

[Fletcher *et al.*, 2019] Justin Fletcher, Ian McQuaid, Peter Thomas, Jeremiah Sanders, and Greg Martin. Feature-based satellite detection using convolutional neural networks. In *Proceedings of the Advanced Maui Optical and Space Surveillance Technologies Conference*, page 11, 2019.

[Harris *et al.*, 2021] Toby Harris, Harriet Brettle, Morgane Lecas, Laurence Blacketer, Antoine Carr, Alberto Fernandez, and Andrea Puppa. An exploration of opportunities to advance ground-based and space-based ssa systems through in-orbit demonstration missions. *ESA Space Debris Office*, 2021.

[Hein, 2020] Guenter W Hein. Status, perspectives and trends of satellite navigation. *Satellite Navigation*, 1(1):22, 2020.

[IAU, 2021] SOFA Board IAU. Iau sofa software collection, 2021.

[Jia *et al.*, 2024] Qianlei Jia, Jiaping Xiao, Lu Bai, Yuhang Zhang, Rangya Zhang, and Mir Feroskhan. Space situational awareness systems: Bridging traditional methods and artificial intelligence. *Acta Astronautica*, 2024.

[Jocher *et al.*, 2023] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[Jocher, 2020] Glenn Jocher. YOLOv5 by Ultralytics, May 2020.

[Kisantal *et al.*, 2020] Mate Kisantal, Sumant Sharma, Tae Ha Park, Dario Izzo, Marcus Märtens, and Simone D'Amico. Satellite pose estimation challenge: Dataset, competition design, and results. *IEEE Transactions on Aerospace and Electronic Systems*, 56(5):4083–4098, 2020.

[Musallam *et al.*, 2021a] Mohamed Adel Musallam, Vincent Gaudilliere, Enjie Ghorbel, Kassem Al Ismaeil, Marcos Damian Perez, Michel Poucet, and Djamila Aouada. Spacecraft recognition leveraging knowledge of space environment: simulator, dataset, competition design and analysis. In *2021 IEEE International Conference on Image Processing Challenges (ICIPC)*, pages 11–15. IEEE, 2021.

[Musallam *et al.*, 2021b] Mohamed Adel Musallam, Vincent Gaudilliere, Enjie Ghorbel, Kassem Al Ismaeil, Marcos Damian Perez, Michel Poucet, and Djamila Aouada. Spacecraft recognition leveraging knowledge of space environment: simulator, dataset, competition design and analysis. In *2021 IEEE International Conference on Image Processing Challenges (ICIPC)*, pages 11–15. IEEE, 2021.

[Park *et al.*, 2023] Tae Ha Park, Marcus Märtens, Mohsi Jawaid, Zi Wang, Bo Chen, Tat-Jun Chin, Dario Izzo, and Simone D'Amico. Satellite pose estimation competition 2021: Results and analyses. *Acta Astronautica*, 204:640–665, 2023.

[Pauly *et al.*, 2023] Leo Pauly, Wassim Rharbaoui, Carl Shneider, Arunkumar Rathinam, Vincent Gaudillière, and Djamila Aouada. A survey on deep learning-based monocular spacecraft pose estimation: Current state, limitations and prospects. *Acta Astronautica*, 2023.

[Peli, 1990] Eli Peli. Contrast in complex images. *JOSA A*, 7(10):2032–2040, 1990.

[Pisutsin *et al.*, 2024] Phumrapee Pisutsin, Jiaping Xiao, and Mir Feroskhan. Omnidrone-det: Omnidirectional 3d drone detection in flight. In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, pages 2409–2414, 2024.

[Proença and Gao, 2020] Pedro F Proença and Yang Gao. Deep learning for spacecraft pose estimation from photo-realistic rendering. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6007–6013. IEEE, 2020.

[Qiao *et al.*, 2022] Sijia Qiao, Haopeng Zhang, Gang Meng, Meng An, Fengying Xie, and Zhiguo Jiang. Deep-learning-based satellite relative pose estimation using monocular optical images and 3d structural information. *Aerospace*, 9(12):768, 2022.

[Redmon, 2018] Joseph Redmon. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[Ren *et al.*, 2016] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time

object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149, 2016.

[Shen *et al.*, 2024] Naijun Shen, Rui Xv, Yang Gao, Chen Qian, and Qingwei Chen. An improved yolov5 model based on feature fusion and attention mechanism for multi-scale satellite recognition. *IEEE Sensors Journal*, 2024.

[Suthakar *et al.*, 2023] Vithurshan Suthakar, Aiden Alexander Sanvido, Randa Qashoa, and Regina SK Lee. Comparative analysis of resident space object (rso) detection methods. *Sensors*, 23(24):9668, 2023.

[Tang *et al.*, 2023] Qiang Tang, Xiangwei Li, Meilin Xie, and Jialiang Zhen. Intelligent space object detection driven by data from space objects. *Applied Sciences*, 14(1):333, 2023.

[Thorsteinson, 2018] Stefan Thorsteinson. Key findings from the neossat space-based ssa microsatellite mission. In *Proc. of the Advanced Maui Optical and Space Surveillance Technologies Conference*, pages 89–103, 2018.

[Usovik, 2023] IV Usovik. Review of perspective space debris mitigation solutions. *Journal of Space Safety Engineering*, 10(1):55–58, 2023.

[Vallado and Crawford, 2008] David Vallado and Paul Crawford. Sgp4 orbit determination. In *AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, page 6770, 2008.

[Vallado, 2001] David A Vallado. *Fundamentals of astrodynamics and applications*, volume 12. Springer Science & Business Media, 2001.

[Wang *et al.*, 2022] Beichao Wang, Shuang Li, Jinzhen Mu, Xiaolong Hao, Wenshan Zhu, and Jiaqian Hu. Research advancements in key technologies for space-based situational awareness. *Space: Science & Technology*, 2022.

[Wang *et al.*, 2024] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.

[Xiao *et al.*, 2024] Jiaping Xiao, Jian Hui Chee, and Mir Feroskhan. Real-time multi-drone detection and tracking for pursuit-evasion with parameter search. *IEEE Transactions on Intelligent Vehicles*, pages 1–11, 2024.

[Zhang *et al.*, 2010] H Zhang, Z Liu, Z Jiang, M An, and D Zhao. Buaa-sid1. 0 space object image dataset. *Spacecraft recovery & remote sensing*, 31(4):65–71, 2010.

[Zhang *et al.*, 2022a] Hao Zhang, Feng Li, Shilong Liu, Lei Zhang, Hang Su, Jun Zhu, Lionel M Ni, and Heung-Yeung Shum. Dino: Detr with improved denoising anchor boxes for end-to-end object detection. *arXiv preprint arXiv:2203.03605*, 2022.

[Zhang *et al.*, 2022b] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu, and Xinggang Wang. Bytetrack: Multi-object tracking by associating every detection box. In *European conference on computer vision*, pages 1–21. Springer, 2022.

[Zhang *et al.*, 2022c] Zipeng Zhang, Chenwei Deng, and Zhiyuan Deng. A diverse space target dataset with multidebris and realistic on-orbit environment. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:9102–9114, 2022.

[Zhang, 2000] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000.

[Zhu *et al.*, 2020] Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, and Jifeng Dai. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*, 2020.