Exploring AI-Based Satellite Pose Estimation: from Novel Synthetic Dataset to In-Depth Performance Evaluation

Fabien Gallet, Christophe Marabotto, Thomas Chambon
Institut de Recherche Technologique Saint Exupéry[*]
{fabien.gallet, christophe.marabotto, thomas.chambon}@irt-saintexupery.com

Abstract

Vision-based pose estimation using deep learning offers a promising cost effective and versatile solution for relative satellite navigation purposes. Using such a solution in closed loop to control spacecraft position is challenging from validation and performance verification viewpoint, because of the complex specification and development process. The validation task entails bridging the gap between the dataset and real-world data. In particular, modelling of Sun power and spectrum, Earth albedo, and atmospheric absence effects, is costly to replicate on ground. This article suggests a novel approach to produce synthetic space scene images. Fine statistical balancing is ensured to train and assess pose estimation solutions. A physically based camera model is used. Synthetic images incorporate realistic light flux, radiometric properties, and texture scatterings. The dataset comprises 120000 images supplemented with masks, distance maps, celestial body positions, and precise camera parameters (dataset publicly available https://www.irt-saintexupery.com/space_rendezvous/ *created in the frame of a project called RAPTOR: Robotic and Artificial intelligence Processing Test On Representative target). An analysis method using a dedicated metric library has been developed to help the assessment of the solution performance and robustness. A deeper comprehension of algorithm behavior through distribution law fitting and outlier identification is then facilitated. Finally, it is shown that implementing Region-of-Interest (RoI) training can drastically increase the performance of the Convolutional Neural Networks (CNNs) for long-range satellite pose estimation tasks.*

1. Introduction

Monocular pose estimation for space rendezvous addresses a significant challenge within the space industry, where the avoidance of expensive LiDAR sensors could simplify access to technologies such as deorbiting, refueling, and in-orbit assembly. However, leveraging visible cameras for this task introduces complex engineering problems, particularly regarding Deep Learning applied to Computer Vision.

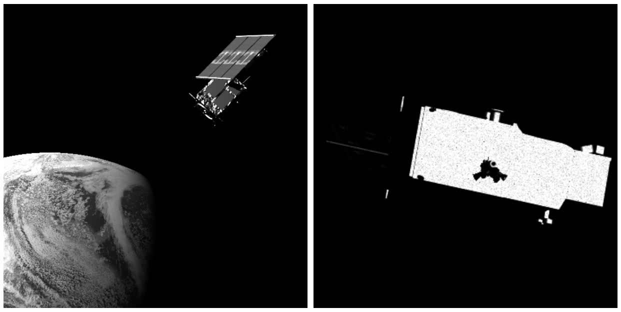
(a) D1 sample, nominal exposure. (b) D2 sample, overexposure.

Figure 1. Examples of RAPTOR dataset images.

CNNs have shown a strong predominance in pose estimation approaches. However, these models are data-intensive and sensitive to domain shifts. In the context of space operations, gathering a vast amount of real-world data is prohibitively expensive. Thus, this lack of data must induce new validation techniques, both for the constitution of more representative synthetic datasets and for the correct quantification of the performances of the proposed architecture.

It is tempting to address the domain gap between training and mission data using a test bench. While the Hardware In the Loop (HIL) solution remains an important part of the functional validation process, space scenes are very difficult to reproduce on ground. The Sun light for instance can be viewed as a very strong punctual source releasing $\approx 1400 W/m^2$ [10] of light to Earth orbiting satellite. The necessary light power is difficult to obtain, in particular because the source must be small or far from the model. A collimated light can be used to overcome this problem but this solution is more suitable for small targets [6].The atmospheric diffusion also brings significant disturbances to the captured scenes. Moreover, the camera is impacted by the orbital environment and the noise level recorded on the ground is often low compared to that in flight [3]. Nevertheless, HIL testing remains useful for real-time constraints and to demonstrate the radiometric validity of the simulator.

The authors propose to evaluate the domain gap using

[*] This work was carried out as part of the RAPTOR project of the French Technological Research Institute *IRT Saint Exupéry*. The authors would like to thank Thales Alenia Space as industrial partner of the project.

simulated images only by mastering key physical and mission properties. The test dataset is divided in two parts, one with typical expected characteristics and another with extended broadcast ranges corresponding to the worst cases that might be encountered in flight conditions.

In summary, the contributions of this article are the following:

1. A 120000 samples dataset specifically designed for spacecraft pose estimation ensuring fine statistical balancing and realistic dynamic ranges using a complete camera model. The camera model contains a physically based ray-tracer coupled with a complete sensor model taking into account key camera limitations. An additional test dataset is provided for testing the robustness of the proposed solution. The provision of data like distance maps, masks and the complete scene description allow to go deeper into the design of pose estimation solutions.
2. The development of a metric library on which to lean a methodology able to better understanding the algorithm to embed it in an on-board closed loop architecture.
3. The performance analysis of a multi-task CNN for long-range pose estimation using RoI training.

2. Related Work

Space applications, especially those involving on-orbit rendezvous, present new challenges for data-intensive supervised learning models. To address these challenges, it is necessary to generate synthetic datasets. Some efforts have been made to provide such datasets to the research community.

For example, in 2019, the Spacecraft Pose Estimation Dataset (SPEED) [9] was introduced as part of the Satellite Pose Estimation Challenge (SPEC2019). This initiative aimed to estimate the pose (i.e., relative position and attitude) of the Tango spacecraft from the PRISMA mission, using individual grayscale images. Subsequently, in 2021, the SPEED+ [14] dataset was introduced, accompanied by a new competition under the Satellite Pose Estimation Challenge (SPEC2021). This update improves the measurement of domain gap robustness with HIL images, sparking interest from the research community. Other datasets such as [2] show a genuine interest in generating high-quality spaceborne synthetic dataset and highlight the difficulties encountered in physically-based modelling. Additionally, datasets like URSO [16] have further analyzed other satellites, such as Soyuz and Dragon. On the other hand, SPARK dataset [12] explores multi-modal spacecraft classification using RGB and depth sensors for a wider range of objects, including 11 classes of spacecraft and debris under various lighting conditions.

Lastly, in 2022, the SEENIC [8] dataset has emerged as a resource suitable for missions involving alternative sensor types, like event cameras. Notably, SEENIC emphasizes the importance of robustness by using a training dataset consisting of simulated event-frames and a testing dataset consisting of HIL data. Unlike its counterparts, SEENIC emphasizes the need for image sequences rather than independant, uniformly distributed scenes.

Deep Learning architectures for pose estimation have evolved considerably in recent years. Two different types of architectures have emerged: direct approaches, which directly predict the orientation and translation of the target object, and indirect approaches, where various representations (such as Heatmaps [4], Part Affinity Fields [5], Dense Maps [18, 22], etc.) are employed to improve network performance. Typically, these representations are post-processed using a variant of the Perspective-n-Point (PnP) [7] algorithm.

In order to assess the performance of pose estimation, conventional metrics typically involve computing the distribution of errors in translation L2-norm and rotation (angle). In [15], for the SPEC challenge requirements, the authors proposed averaging errors across the dataset and introduced a global metric as the sum of mean errors in rotation and translation. While these metrics facilitated comparison of solutions within the challenge framework, they lacked the ability to quantify the extent of error distribution or its dependency on distance. However, for space rendezvous, a clear understanding of how performance improves relative to distance is crucial to ensure trajectory convergence and mitigate collision risks. Furthermore, for operational deployment in closed-loop control systems, a commitment to distribution quantiles is necessary.

3. Dataset Generation

3.1. SPICaM Overview

SPICaM, for Spacecraft and Planetary Imaging by Camera Modeling, is a complete camera model composed of a 3D rendering engine and a sensor model. The rendering engine is able to perform physically based computation of radiance maps. It uses a ray-tracing mechanism associated to energy conservative BRDF (for Bidirectional Reflectance Distribution Function) models in order to estimate realistic light fluxes. Contrary to most known commercial rendering engines which are focused on human perception, SPICaM allows to compute radiances on custom bands of light spectrum considering respective sensor efficiency. Light fluxes on each band are converted to electrons with a given exposure time. The electrons are then converted to bits, each conversion comes with injection of corresponding disturbances (optical scattering, distortions and noise, DSNU (Dark Signal Non-Uniformity), PRNU (Photo-Response Non-Uniformity), dark current, readout noise).

3.2. Simplified Camera Model

A simplified camera model has been designed to predict how much light flux will be received per satellite face. The client satellite is modeled here as six rectangular faces for the body and two faces for the solar array. Each face has its own optical response and one light ray is traced by client face. Such approach could be embarked on board and allows to approximate the best exposure time for each scene. The best exposure time is defined as the duration of integration of the light flux which maximizes the quantity of usable information in the image. The satellite is therefore always visible with a good level of detail in each image.

3.3. Dataset Domains

In order to model a representative domain gap, we have chosen to differentiate nominal (D1) and perturbed (D2) domains based synthetically on optical, sensor, scene and target parameterization. These datasets allow to measure performance over a first level of realistic perturbations. Training, validation and test split are released on D1 domain. A second test dataset on D2 domain. Tab. 2 defines the varying parameter ranges of each domain.

	Domain	
Split	Nominal (D1)	Perturbed (D2)
Train	64000	-
Validation	16000	-
Test	20000	20000

Table 1. Dataset composition per domain and split.

3.4. Statistical Distributions

The dataset is designed to maximize the uniformity of the distribution for the varying parameters while guaranteeing situations respecting the physical laws. A filtering of invalid samples is performed all along the process (see Section 3.5).

3.4.1 Main Assumptions

The following assumptions were made to define the dataset.

- A single orbit corresponding to one given fictive mission is considered.
- The scattering methods used are valid for any Earth centered orbit.
- It is modelled a single client corresponding to Sentinel-3 satellite.
- The dataset is optimized for keypoint-based pose estimation solutions (the process tries to enhance uniformity of the distributions of keypoints positions in the image).
- Optimal exposure estimation is performed on-board (it appears mandatory for the chaser to be able to estimate the camera exposure hence allowing to see the client clearly). The section 3.2 explains the associated method.
- The camera parameters selected are: 4 Mpix 8bits CMOS monochromatic sensor, 60° of field of view.

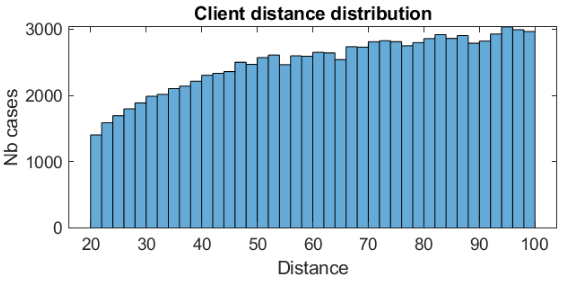

Figure 2. Distribution of relative distance between the client and the servicer (D1 dataset).

3.4.2 Scattered Parameters

The dataset is associated to a fictive but realistic mission with a client spacecraft in GEO (Geosynchronous Earth Orbit). The samples are uniformly scattered over a one year range to cover the full Earth orbit period corresponding to various Earth backgrounds.

The attitude of the satellite is therefore defined by a random quaternion in order to encompass all possible configurations that could correspond to breakdown situations. During the mission, the chaser can be anywhere with respect to the client satellite reference frame. The pose estimation solution is designed to be efficient over a range of distances. A spacecraft that is too close may not have all of its recognizable features within the camera's field of view. The range is also limited by the minimum amount of pixels imaging the spacecraft. As illustrated in Fig. 2, the distance between the spacecrafts is drawn using a uniform distribution but the closer the client is, the more difficult it is to place him entirely inside the frame, thus changing the distribution.

The target to chaser direction is drawn randomly in order to obtain a uniform distribution on the surface of the unity sphere. Drawing the chaser's attitude directly would have led to numerous instances of the client being out of frame. Consequently, the client's position within the image is determined using random pixel coordinates. This approach facilitates the establishment of an initial direction in 3D space to define the chaser's attitude. As depicted in Fig. 3, placing the satellite entirely within the image becomes challenging when its midpoint is near the edges of the image. Fig. 4 complements the previous one by presenting the distributions of keypoints positions within the image. the probability of having centered keypoints is higher when the satellite is close for the proposed configuration. This explains the drop in probability at the edges of the image. The chaser attitude is deduced from the client satellite relative position

Table 2. Varying ranges between D1 and D2 domains.

Parameter	Type	D1	D2
PSF fwhm Distance	Optical	0.4 pix	0.4 pix ± 0.1
Exposure time	Sensor	optimal exposure = t	$t \times 0.5$ to 2
PRNU	Sensor	sigma = 1/100	sigma = 2/100
MLI texture	Target	1 texture	5 textures
Material BRDF	Target	1 fixed BRDF	Albedo coefficient × 0.5 to 1

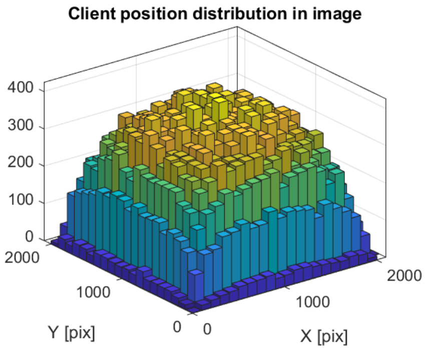

Figure 3. Distribution of the client position in the image.

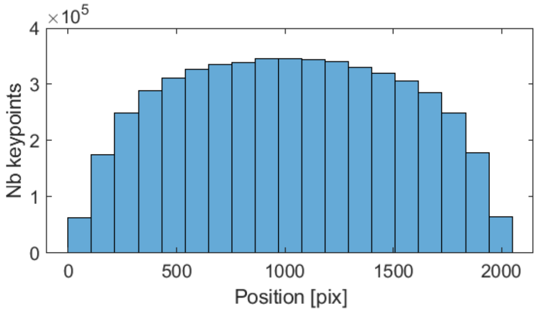

Figure 4. Distribution of all keypoints positions (X/Y) in images.

and its projected position on sensor frame. The roll around client direction axis is uniformly drawn between $-\pi$ and π.

A nominal exposure duration is computed using a simplified camera model (see dedicated Sec. 3.2). The nominal duration is only scattered in D2 test dataset by applying a random coefficient uniformly distributed between 0.5 and 2. Fig. 1a and Fig. 1b show two samples with nominal and scattered exposure.

The Multi-Layer Insulation (MLI) covers the majority of the selected client. This material is flexible and usually has many unpredictable wrinkles. This is modeled using a normal map, drawn inside a set of five reference textures. The optical properties of materials can only be partially known due to the exposure to space environment but also the lack of design data on the subject. The client body cover and structure is mainly composed of MLI and aluminum. The optical response of the materials is modeled using a GGX BRDF model [21]. The albedo coefficient is scattered between the nominal value and half of it over each spectral band in order to simulate the material clouding.

The sensor usually has several defects; some of it possibly corrected after a calibration step. We consider that the dark current can be partly calibrated and that the rest is not significant (less than one bit for our range of exposure duration). The read-out noise is typically very faint ($<$ 2 bits with defined camera hypotheses). We identified the Photo-Response Non-Uniformity (PRNU) as the potentially most impacting noise for pose estimation. It is firstly a multiplicative noise hence introducing a non-linear transformation. The PRNU is also hard to calibrate and intended to be degraded over satellite life-time.

3.5. Automatic Sample Filtering

Several checks are performed all along the scattering process in order to automatically discard low quality data.

- **Earth Eclipse**
  The client in GEO orbit is not visible without direct light. A conical shadowing model is used to detect eclipses and remove these cases.
- **Camera Blinding**
  A camera is qualified to be used considering a Sun exclusion angle. The dataset is therefore constructed considering that this constraint is fully respected during the rendezvous mission.
- **Keypoints Outside of Frame**
  The system verifies the presence of keypoints beyond the camera frame to ensure comprehensive coverage of the entire satellite within the captured image.
- **Client Visibility**
  Utilizing the simplified camera model, an estimation of client visibility is derived, with recognition of the significant impact on visibility of the Sun incidence and the client attitude. Scenes lacking visible faces are promptly discarded.

3.6. Additional Meta-Data

The dataset is provided with additional meta-data intended to open the door to further solutions. Additional masks available for segmentation purposes are provided as 8 bits mono-band images. The pixel values correspond to the scene element intersected by the light ray (Earth, Moon, client body or client solar array). Distance maps are also provided as well as a complete definition of the scene, paving the way for the design of other image processing architectures.

4. Analysis Methodology and Metric Library

To compare and analyze efficiently the results provided by various pose restitution approaches, an analysis methodology for performance computation has been set up. The main objectives are the following:

1. Proposing a methodology to analyze pose estimation performances and able to compare solutions and monitor learning processes;
2. Quantifying performance distributions, and not only average behaviour;
3. Understanding pose performance variations with respect to target distance;
4. Quantifying the robustness of the solution, by estimating the ratio of outliers in the outputs.

Robustness quantification is of importance for the integration of pose estimation algorithm in closed loop systems. The objective is to quantify the level of outliers to be filtered out by navigation algorithms.

4.1. Optimal Standard Deviation Law

In the proposed approach, performance metrics are computed in the chaser camera frame, and rotations are represented by rotation angles and axis vector 3D (dimension 3) representation. The errors therefore lie in a 6D (dimension 6) domain and are defined as follows for a given data point:

$$e_r = \hat{r}_{cf} - r_{cf} \tag{1}$$

$$e_q = rotvec\left(|\langle \hat{q}_{cf}, q_{cf}^T \rangle|\right) \tag{2}$$

where

- index cf means *camera frame*,
- $\langle \cdot, \cdot \rangle$ is the quaternion inner product notation,
- $rotvec$ is the conversion function from quaternion to angle parametrization,
- r_{cf} and $\hat{r}_{cf}$ represent the true and estimated positions in camera frame,
- q_{cf} and $\hat{q}_{cf}$ represent the true and estimated unit quaternions.

One of the objectives of the metrics model is to better understand the variation of the pose estimation performances with respect to distance z from the target. Such errors typically demonstrate a quadratic variation with respect to the distance. This law depends on the geometrical definition of the problem, e.g., dependence on the object shape, on the angular field of view of the camera and its resolution.

To represent the distance variation, a standard deviation parametrization depending on z^2 is proposed:

$$\tilde{\sigma}(z) = a + bz + cz^2 \tag{3}$$

This law is assumed independent for each one of the 6D components of the error vector. To estimate the parameters of the law, the likelihood estimator is maximized. Under the Gaussian assumption, such an estimator can be written for a given sample population X with errors x_i as follows:

$$P(X) = \prod_{i=1}^{N} \frac{1}{\tilde{\sigma}(z_i)} e^{-\frac{1}{2}\left(\frac{x_i}{\tilde{\sigma}(z_i)}\right)^2} \tag{4}$$

One can rewrite the negative likelihood log F to be minimized as follows:

$$\begin{aligned} F(a,b,c) &= -\frac{1}{N} log(P) \\ &= \frac{1}{N}\sum_{i=1}^{N} log(\tilde{\sigma}(z_i)) + \frac{1}{2N}\sum_{i=1}^{N} \frac{x_i^2}{\tilde{\sigma}(z_i)^2} \end{aligned} \tag{5}$$

This function is not convex, but depends on the inverse of a square when close to zero and as a logarithm at longer range, especially if data are well distributed over z. Finding its minimum can be solved typically via Gauss-Newton algorithms using efficient computation of gradient and Hessian of the function:

$$\nabla(F) = \begin{bmatrix} \frac{\partial F}{\partial a} \\ \frac{\partial F}{\partial b} \\ \frac{\partial F}{\partial c} \end{bmatrix} = \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} 1 \\ z_i \\ z_i^2 \end{bmatrix} \left(\frac{1}{\tilde{\sigma}(z_i)} - \sum_{i=1}^{N} \frac{x_i^2}{\tilde{\sigma}(z_i)^3} \right) \tag{6}$$

$$H(F) = \frac{1}{N}\sum_{i=1}^{N} \begin{bmatrix} 1 & z_i & z_i^2 \\ z_i & z_i^2 & z_i^3 \\ z_i^2 & z_i^3 & z_i^4 \end{bmatrix} \left(-\frac{1}{\tilde{\sigma}(z_i)^2} + 3\sum_{i=1}^{N} \frac{x_i^2}{\tilde{\sigma}(z_i)^4} \right) \tag{7}$$

4.2. Outliers and Distribution Estimation

Outliers are estimated based on Gaussian assumption. Such an assumption is motivated by the fact that many navigation filter architectures rely on the hypothesis of Gaussian distributions to find the optimal solution.

The outliers are estimated directly in 6D, through the following an algorithm composed of the following steps:

1. Normalization of errors taking into account distance model presented previously.

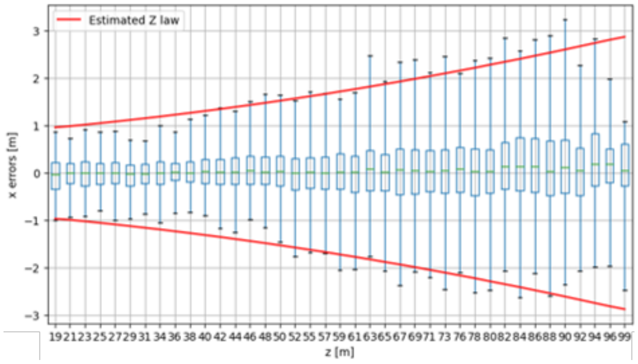

Figure 5. Example Z law fitting for translation errors. Blues boxes: the error distribution (up to 90% quantiles). Red curve: the fitted 90% quantile law.

2. Robust estimation of data covariance (an algorithm described in [17] is used).
3. Estimation of outliers based on the Mahalanobis' distance computed as:

$$D_M(x) = \sqrt{(x-\mu)^T\,\Sigma^{-1}\,(x-\mu)} \tag{8}$$

where x represent a generic data point, μ is the distribution mean, and Σ is the covariance matrix.

Assuming Gaussian distributions in 6D, the square of Mahalanobis' distribution should follow a law in 6D. Outliers are computed as points whose errors are greater than the Mahalanobis' theoretical quantile that would gather 99% of the 6D distribution.

4.3. Performance Metrics

Performance metrics are defined as hold over 90% of the dataset. Since performances differ significantly depending on the axes, they are computed in 3D both in the translation and rotation domains. They correspond to the ellipsoid that would gather 90% of the data. The size of the ellipsoid can be computed based on the 90% quantile of the Mahalanobis' distance distribution.

The final performance metrics is then characterized by the ellipsoid inner volume represented by the equivalent sphere radius R_s (Eq. (9)).

$$R_s = Q_M(0.9)\det(\Sigma)^{(1/3)} \tag{9}$$

where $Q_M(0.9)$ corresponds to the 90% quantile over the Mahalanobis' distance distribution, and $\det(\Sigma)$ is the determinant of the covariance matrix.

5. Experiments

This section is dedicated to demonstrating the good quality of the proposed dataset, primarily through the utilization of the proposed metrics on a reference pose estimation architecture.

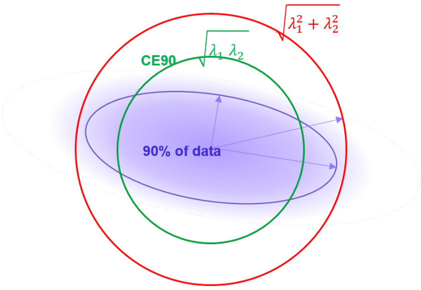
$\sqrt{\lambda_1^2+\lambda_2^2}$
CE90
$\sqrt{\lambda_1\,\lambda_2}$
90% of data

Figure 6. Definition of the CE90 metric related to the equivalent ellipsoid volume.

5.1. Pose Estimation Architecture

Current pose estimation techniques can be categorized in two types of approaches: direct approaches (predicting directly orientation and relative position of the target object) and indirect approaches (predicting intermediate representations). In this section, results obtained with a hybrid *multi-task* approach called SPNv2 are presented. SPNv2 approach [13] has been developed at the Space Rendezvous Laboratory (SLAB) from Stanford University. This model is based on an EfficientNet [19] backbone, a multi-scale feature fusion step called weighted Bi-directional Feature Pyramid Network (BiFPN) [20] directly plugged to multiple prediction heads. This model simultaneously performs the following tasks:

1. Classification: binary classification using α-balanced variant of Focal Loss [11].
2. Object Detection: prediction of one bounding box per object (xmin, ymin, xmax, ymax) using Complete Intersection-over-Union (CIoU) loss [23] .
3. Direct Pose Estimation: rotation regression using a 6D representation [24] and translation regression using SPEED score [9]. With $\tilde{R}$ and R (i.e., $\tilde{t}$ and t) respectively corresponding to the rotation matrix (i.e., translation vector) predicted and ground truth, final pose error E_{pose} is defined as follows:

$$E_{pose} = E_R(\tilde{R}, R) + E_T(\tilde{t}, t)/\|t\| \tag{10}$$

with

$$E_R(\tilde{R}, R) = \frac{arccos(tr(R^T, \tilde{R}) - 1}{2} \tag{11}$$

and

$$E_T(\tilde{t}, t) = \|\tilde{t} - t\| \tag{12}$$

4. Keypoint Detection: keypoint-wise heatmap for keypoint detection using pixel-wise Mean Square Error (MSE) between the predicted heatmaps $\tilde{h}$ and the ground truth heatmaps h. Pose is then retrieved using an algorithm such as Perspective-n-Point (PnP) from a 3D reference model and camera parameters.

$$MSE(\tilde{h}, h) = \sum_{i=1}^{n} (\tilde{h_i} - h_i)^2 \qquad (13)$$

5. Binary Semantic Segmentation: differentiation of the pixels related to the detected objects from the background using pixel-wise binary cross entropy on logits.

Originally, all losses are optimized simultaneously with equal weights. Nonetheless, we choose to not perform binary semantic segmentation as this head offers no significant performance gains [13]. Object detection have also been deactivated when RoI training is applied. For this experiment, we keep original data augmentation (brightness and contrast, sun flare, blur, noise and random erase) proposed in SPNv2 [13]. In order to improve long-range performance, the impact of an imperfect object detector on system performance is studied using RoI training and bounding box data augmentation.

5.2. Region-of-Interest Training

Crop functions reduce image size at the entrance of the pose estimator while maintaining content information as much as possible. When operated in closed loop, the crop is defined based on the current estimate of target position in the full resolution image. Depending on the cases, the resizing function performs a down-sampling or an up-sampling step. Down-sampling is the most demanding operation while aliasing must be contained to limit noise at the entrance of the pose estimation network.

When applying crop, the camera projection function K from the projective equation is changed through a pixel wise translation and a zoom factor. The coordinates of a 3D point M_i in the image plane depend on the rotation matrix R and translation vector t from object frame to camera frame, the 3D point itself M_i, the camera matrix K:

$$m_i = m_i(R, t, M_i, K) \qquad (14)$$

The transformation corresponding to zoom and translation in the image plane cannot be represented in the pose domain of full resolution image. This change is straightforward for keypoints that at first order can be considered as being translated by neglecting distortion. However, for what regards direct pose estimation, the problem is more complex. The implemented solution consists in computing the closest pose solution that minimizes the re projection errors in image coordinates.

5.3. Results

After analysing of the output performances on the D1 and D2 test datasets, the following main conclusions can be derived.

1. The outputs confirm the quadratic dependence of the performances with respect to distance (see Fig. 8).
2. Performances of the direct head are close to be Gaussian while PnP has more elongated tail. Direct heads have therefore less outliers (see Fig. 7 and Tab. 3). This result is explained by matrix inversions occurring in PnP approaches. Indeed, even if keypoints errors are close to be Gaussian, their contribution is part of matrix inversion in SQPnP algorithm.
3. In-plane performances (rotation abound Z axis and translation along X and Y axes) are far better than across plane performances. This behaviour is an inherent limitation of single camera pose estimation. Moreover, the results show that the errors in the Z direction are correlated with X and Y axes.
4. The implementation of RoI training enables us to achieve better performances (see Tab. 3).
5. An expected degradation can be observed on D2 with respect to D1, without drastically changing the amount of outliers.
6. Systematic convergence of the PnP algorithm over D1 and D2 confirming the good quality of the dataset.

6. Discussion

The proposed dataset offers a key contribution by using a physical renderer and implementing a fine management of material properties as well as space environment.

The actual rendering engine still suffers from few limitations. The effect of parasitic light inside the camera optical system is not yet modeled. While a careful definition of the approach strategy can avoid disturbing Sun incidences, minor stray-light and flares would be observed in flight. In addition, motion blur has been neglected from now on. This effect might have an effect when relative dynamics are higher or camera line-of-sight shaken due to client satellite control. Theses two features will be implemented in future rendering engine releases.

Real images taken on the ground are always interesting to assess robustness to domain gap. Even if representative space is generally limited, testing on a robotic test bench is clearly a mandatory step to validate a pose estimation solution. This implies ensuring that the solution is fully functional on this means of testing and therefore robust to the corresponding change of domain. It is planned to complement the existing dataset with a subset of real images of space scenes produced using a robotic testbed.

Data augmentation and domain adaptation techniques could be improved to limit the performance degradation in disturbed conditions. Moreover, the addition of numerous metadata allows to investigate deeper into multi-tasking and multi-modal architectures.

The performance metrics proposed enable a deeper comprehension of error distributions and facilitate the automatic control the architecture outcomes in relation with operational needs. Some limitations can still be discussed such

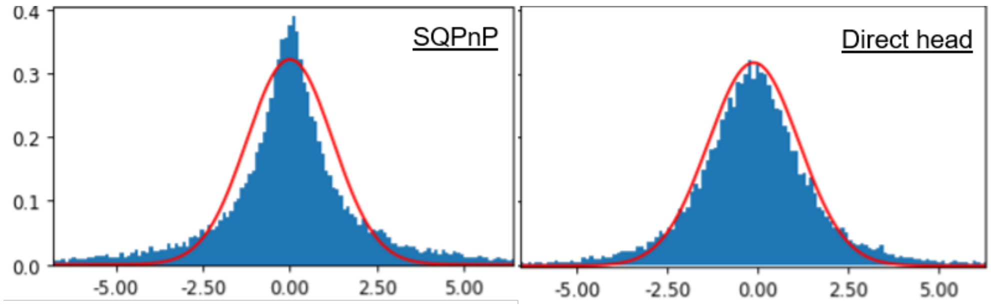

Figure 7. Illustration of typical translation distribution errors. Left: SQPnP results. Right: direct pose estimation head.

Table 3. Performances over test datasets from SPNv2-B0 keypoint detection and direct pose estimation heads.

		Keypoint Detection head (SQPnP)			Direct Pose Estimation head		
Domain	Architecture	Outliers	q (CE90) [°]	r (CE90) [m]	Outliers	q (CE90) [°]	r (CE90) [m]
1	SPNv2-B0	30%	60	6.0	20%	50	4.5
1	SPNv2-B0 (RoI)	30%	7	1.2	15%	14	2.2
2	SPNv2-B0	31%	78	7.6	23%	64	5.6
2	SPNv2-B0 (RoI)	30%	10	2.1	17%	16	3.0

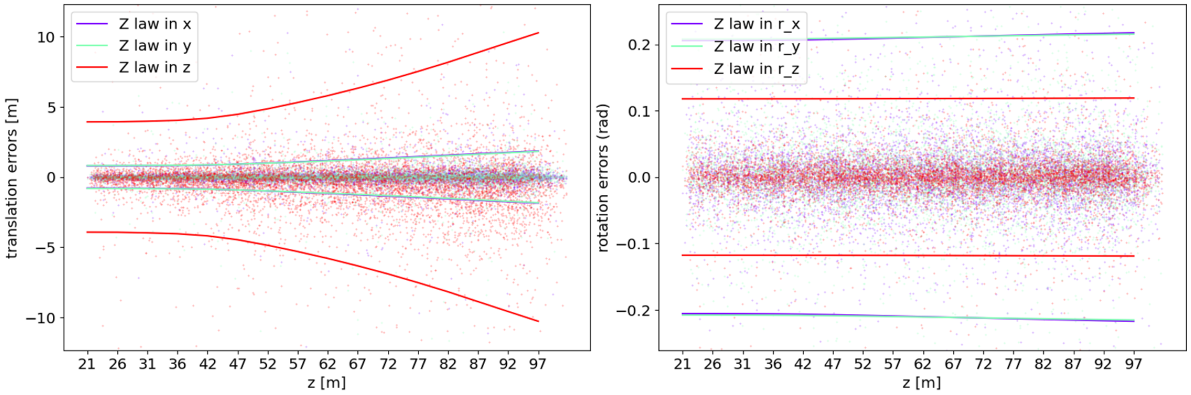

Figure 8. CE90 estimated law over domain D2 with respect to distance to the target (with RoI). Left: translation. Right: rotation. Each curve gives the ellipsoid radius along a given axis. The point cloud corresponds to the actual performance data zone.

as the hypothesis of Gaussian distributions used to estimate the distance model. In case of the use of PnP algorithms, we actually know that the distributions are not Gaussian and a dedicated model with longer tail laws (such as Student's law or $sinh$ laws [1]) could improve the results as well as being more generic and robust.

7. Conclusions

In this work, a new pose estimation dataset is proposed. The dataset is associated with a specific methodology of analysis based on a dedicated metric library. The dataset is built using a unique camera model conceived to be representative to satellite flight cameras. Scenes are defined to conform to a typical GEO orbit rendezvous scenarios, while minimizing distribution biases. The produced main dataset of 100000 images (including mask, distance maps and scenes definitions) is abounded with another 20000 images test dataset modelling realistic worst-case situations.

The overall dataset is used to train and test various pose estimation architectures. An analysis of a typical computer vision solution with RoI training revealed significant improvements for long-range pose estimation.

The proposed metric library allows to define an analysis methodology to accurately quantify the performance and the robustness of the solution. It also offers useful tools to facilitate the customer-supplier interface by helping to construct relevant requirements for such a subsystem.

References

[1] Tensorflow user manual webpage: tfp.distributions.sinharcsinh. https://www.tensorflow.org/probability/api_docs/python/tfp/distributions/SinhArcsinh. 8

[2] Michele Bechini, Paolo Lunghi, Michéle Lavagna, et al. Spacecraft pose estimation via monocular image processing: Dataset generation and validation. In *9th European Conference for Aerospace Sciences (EUCASS 2022)*, pages 1–15, 2022. 2

[3] Heidi Becker. Commercial sensor survey radiation testing progress report. 1

[4] Yannick Bukschat and Marcus Vetter. Efficientpose: An efficient, accurate and scalable end-to-end 6d multi object pose estimation approach. *arXiv preprint arXiv:2011.04307*, 2020. 2

[5] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7291–7299, 2017. 2

[6] Mehregan Dor, Travis Driver, Kenneth Getzandanner, and Panagiotis Tsiotras. Astroslam: Autonomous monocular navigation in the vicinity of a celestial small body – theory and experiments. (arXiv:2212.00350), 2022. arXiv:2212.00350 [cs, eess]. 1

[7] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2

[8] Mohsi Jawaid, Ethan Elms, Yasir Latif, and Tat-Jun Chin. Towards bridging the space domain gap for satellite pose estimation using event sensing. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11866–11873. IEEE, 2023. 2

[9] 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. 2, 6

[10] Greg Kopp and Judith L. Lean. A new, lower value of total solar irradiance: Evidence and climate significance. *Geophysical Research Letters*, 38(1), 2011. 1

[11] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 6

[12] Mohamed Adel Musallam, Kassem Al Ismaeil, Oyebade Oyedotun, Marcos Damian Perez, Michel Poucet, and Djamila Aouada. Spark: spacecraft recognition leveraging knowledge of space environment. *arXiv preprint arXiv:2104.05978*, 2021. 2

[13] Tae Ha Park and Simone D'Amico. Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap. *Advances in Space Research*, 2023. 6, 7

[14] Tae Ha Park, Marcus Märtens, Gurvan Lecuyer, Dario Izzo, and Simone D'Amico. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In *2022 IEEE Aerospace Conference (AERO)*, pages 1–15. IEEE, 2022. 2

[15] 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. 2

[16] Pedro F Proença and Yang Gao. Deep learning for spacecraft pose estimation from photorealistic rendering. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6007–6013. IEEE, 2020. 2

[17] Peter Rousseeuw and Katrien Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999. 6

[18] Yongzhi Su, Mahdi Saleh, Torben Fetzer, Jason Rambach, Nassir Navab, Benjamin Busam, Didier Stricker, and Federico Tombari. Zebrapose: Coarse to fine surface encoding for 6dof object pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6738–6748, 2022. 2

[19] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 6

[20] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. 6

[21] Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. Microfacet models for refraction through rough surfaces. 4

[22] Gu Wang, Fabian Manhardt, Federico Tombari, and Xiangyang Ji. Gdr-net: Geometry-guided direct regression network for monocular 6d object pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16611–16621, 2021. 2

[23] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence*, pages 12993–13000, 2020. 6

[24] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5745–5753, 2019. 6