POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Unseen Non-Cooperative Spacecraft Pose Estimation via Image-Based Learning

Laurea Magistrale in Computer Science and Engineering - Ingegneria Informatica

Author: Lorenzo Bottelli

Advisor: Prof. Michèle Roberta Lavagna

Co-advisor: Mathieu Salzmann

Academic year: 2025-2026

1. Introduction

Pose estimation of a non-cooperative spacecraft plays a crucial role in enabling the deployment of automatic vision-based systems in orbit, with applications ranging from In-Orbit Servicing (IOS) to Active Debris Removal (ADR). A non-cooperative target refers to a space object that lacks dedicated equipment, such as attitude control functions and communication capabilities, thereby increasing the technical complexity of approach, inspection, or capture operations by a servicer spacecraft. These proximity operations require accurate knowledge of the target's 6 Degrees-of-Freedom (6D) pose, namely its relative position and attitude. In recent years, deep learning-based methods demonstrated strong performance in estimating the pose of a known spacecraft from monocular images [5]. Networks were trained and evaluated on single-object datasets such as SPEED+ [4], which limits their ability to generalize to novel, previously unseen targets. More recently, research efforts have addressed the problem of 6D pose estimation of arbitrary unseen objects in terrestrial environments, achieving competitive performance on standard object pose estimation benchmarks. This thesis investigates the application of a state-of-the-art terrestrial unseen object pose estimation framework to spacecraft proximity operations, demonstrating that generalizable deep learning models can be effectively transferred to the space domain. This enables relative navigation with non-cooperative spacecraft without requiring target-specific training. The main contributions of this work include the generation of a multi-target dataset of 10,000 synthetic images comprising 28 spacecraft models from the ESA public repository, the introduction of domain-specific adaptations to the estimation framework, and a comprehensive evaluation of pose estimation performance on previously unseen spacecraft. Furthermore the method was successfully validated on real in-orbit imagery acquired by Astroscale Japan during a fly-around of the JAXA H-2A upper stage [2]. This thesis has been conducted under the supervision of the ASTRA Laboratory at Politecnico di Milano and the CVLab at EPFL.

2. Pose Estimation Method

Given a single image of an object, 6D pose estimation is the problem of finding the rigid 6D transformation from the object to the camera by predicting the relative translation vector **t**

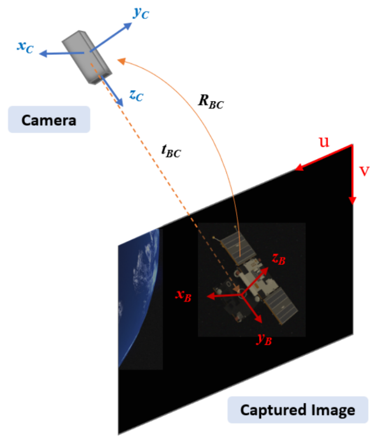

Figure 1: Representation of the 6D pose estimation problem.

and rotation matrix R, as represented in Figure 1. For unseen object pose estimation, two main paradigms have emerged depending on the information available at inference time: *model-based* methods, which assume access to a 3D CAD model of the target object, and *model-free* methods, which rely instead on a set of reference images. When dealing with spacecraft targets, the availability of a 3D CAD model is generally a realistic assumption, whereas acquiring a sufficiently diverse set of representative reference images is significantly more challenging. After reviewing state-of-the-art models, *MegaPose* [3] was selected for three main reasons: it is a *model-based* method, it directly regresses the translation and rotation components of the pose thus avoiding the computational overhead associated with *PnP* geometric solvers, and it relies on a CNN-based architecture, making it more suitable for hardware-constrained environments such as spacecraft on-board systems compared to Vision Transformer-based alternatives. *MegaPose* consists of two main components: a coarse estimator, which provides an initial pose estimate by classifying several pose hypotheses using a render-and-compare strategy, and a refiner, which iteratively improves the coarse estimate. The model was trained on a dataset of 2 million synthetic images of numerous classes of everyday objects, and the metrics used to evaluate performance are primarily designed for robotic and terrestrial applications, where object scales and operating distances differ substantially from those encountered in spacecraft proximity operations. For this reason, the evaluation in this work employs metrics tailored to spacecraft pose estimation. Let $\mathbf{t}_{pred}$ and $\mathbf{t}_{GT}$ be the predicted and ground-truth translation vectors, and $\mathbf{q}_{pred}$ and $\mathbf{q}_{GT}$ the corresponding rotation quaternions. The position, orientation and final pose scores for image i are defined in Equations (1).

$$
\begin{aligned}
\text{score}^{(i)}_{position} &= \frac{||\mathbf{t}^{(i)}_{GT} - \mathbf{t}^{(i)}_{pred}||}{||\mathbf{t}^{(i)}_{GT}||} \\
\text{score}^{(i)}_{orientation} &= 2 \cdot arccos(|\langle \mathbf{q}^{(i)}_{pred}, \mathbf{q}^{(i)}_{GT} \rangle|) \\
\text{score}^{(i)}_{pose} &= \text{score}^{(i)}_{position} + \text{score}^{(i)}_{orientation}
\end{aligned}
\tag{1}
$$

Additional modifications were introduced to adapt the framework to the constraints of spacecraft imagery. In particular, the lack of segmentation masks, normally used by *MegaPose* to discard occluded objects or instances outside the image boundaries, was handled, and rendering parameters used for pose hypothesis generation were adjusted, including the virtual camera far clipping plane and the option to enable textured rendering.

3. Multi-target Dataset

Existing spacecraft pose estimation datasets are all single-target. Consequently, learning-based pose estimation methods trained on such datasets tend to overfit to a specific target and often fail to generalize to previously unseen spacecraft. To address this limitation, a multi-target dataset was generated to encompass a broad range of spacecraft geometries, scales, and surface material properties, as well as diverse viewing and illumination conditions representative of realistic orbital scenarios. The dataset was generated using ADE-SATELLITE, a numerical tool designed for visible-spectrum spacecraft imaging based on the methodology presented in this research [1]. The selected orbital altitude, relative camera-to-target distance, and Sun aspect angle ranges were chosen to reflect typical Low Earth Orbit (LEO) proximity operation scenarios. Object poses were sampled randomly to ensure coverage of a wide variety of orbital configurations. A total of 8,000 training samples and 2,000 testing samples were generated. Half of the images include the Earth as background, while the remaining half de-

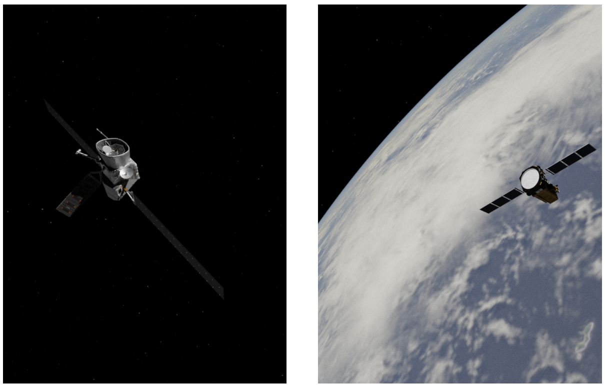

Figure 2: Examples of rendered images.

pict deep-space backgrounds. The 3D spacecraft models were obtained from the ESA public repository, with training and test targets strictly separated to prevent data leakage and to guarantee evaluation on previously unseen objects. Test targets were deliberately selected to ensure diversity in shape, scale, and surface appearance. Figure 2 presents examples of rendered images.
The pose annotations produced by the rendering tool are expressed in the Earth-Centered Inertial (ECI) coordinate frame. These annotations were therefore transformed to obtain the target pose relative to the camera, which is the convention adopted for both training and evaluation. The complete dataset is available at: *Multi-target Spacecraft Pose Estimation Dataset*

4. Experiments, Results and Analysis

This section presents and discusses the main experiments and analyses conducted in this work. In particular, two configurations of the pre-trained model and two configurations of the fine-tuned model are considered.

4.1. Experiments

The evaluation was performed on the test dataset described in Section 3. Predicted poses were matched with the corresponding ground-truth annotations to compute the position, orientation, and overall pose scores defined in Equations 1, where lower values indicate better accuracy. In the first configuration of the pre-trained model, the renderings used by the coarse estimator were generated from plain 3D meshes without texture information. In the second configuration, textured rendering was enabled to provide richer appearance information for pose hypothesis evaluation. The fine-tuning strategy consisted of freezing the backbone network to preserve previously learned generic features while training only the task-specific head of the refiner. As with the pre-trained model, the fine-tuned network was evaluated using both binary and textured renderings.

4.2. Results

Table 1 reports the mean pose scores obtained across the full test set for the four considered configurations.

Model configuration	Pose Score
Pre-trained, binary render	0.9220
Pre-trained, textured render	**0.4226**
Fine-tuned, binary render	1.0052
Fine-tuned, textured render	0.5887

Table 1: Mean pose scores on the test set.

The best performance is achieved by the pre-trained model with textured rendering enabled. Focusing on this configuration, Figures 3 and 4 illustrate the distribution of position and orientation scores, and the per-target mean score breakdown, respectively. Qualitative pose prediction visualizations are presented in Figure 5.

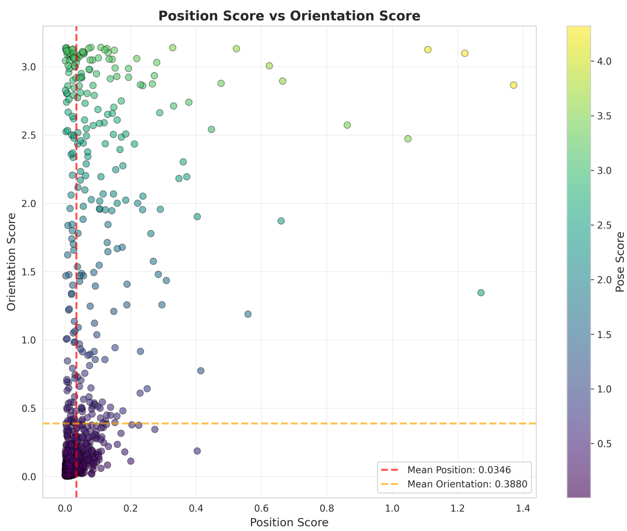

Figure 3: Position vs orientation score distribution.

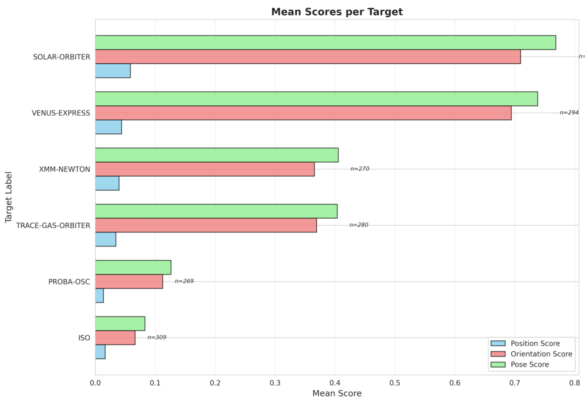

Figure 4: Per-target mean scores.

The superior performance of the pre-trained model suggests that fine-tuning may have introduced overfitting to the synthetic training dataset. This interpretation is supported by the observed reduction in training and validation losses during fine-tuning, coupled with degraded generalization performance on the unseen test set.

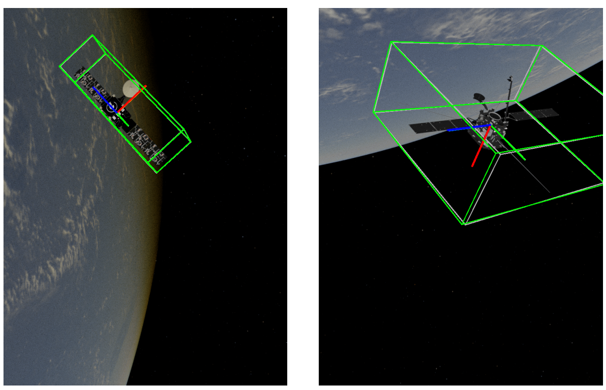

Figure 5: Pose estimation visualizations.

4.3. Analysis

Target symmetry was identified as one of the primary sources of pose estimation errors. To quantify the impact of symmetric ambiguities, an analysis of the orientation error distribution was conducted. In particular, the concentration of orientation errors around 180° was examined, as this corresponds to flipped pose predictions commonly induced by symmetric target geometries. The analysis was performed by fitting a Gaussian Kernel Density Estimator (KDE) to the orientation error distribution and detecting local maxima to identify dominant peaks. Table 2 reports the identified peaks along with the percentage of samples falling within a 10° interval around each peak. The per-target analysis further indicates that spacecraft with strongly asymmetric structures do not exhibit a significant secondary peak, whereas more symmetric targets show a higher proportion of samples clustered around the flipped-orientation peak.

Peak value	Interval	Concentration
3.3°	[0.0, 8.3]°	72.5%
171.9°	[166.9, 176.9]°	2.3%

Table 2: Concentration of orientation errors around peaks identified via KDE.

A complementary analysis was conducted to investigate the impact of background composition on prediction accuracy by separating the results according to background type. The mean pose score for samples containing the Earth in the background is 0.3792, whereas samples with a deep-space background yield a mean score of 0.4662. These results confirm the expected trend: deep-space backgrounds provide lower visual contrast, making the extraction of discriminative object features more challenging.
As previously discussed, the primary factor contributing to performance improvement is the use of textured rendering. This effect was quantitatively assessed, showing that textured rendering reduces the coarse pose error by 34.54%. In the non-textured setting, the refinement stage reduces the pose error by 21.9%, whereas in the textured setting the reduction increases to 45.31%. This demonstrates that the refiner is significantly more effective when initialized with a higher-quality coarse pose estimate.

5. Validation on Real Mission Data

A key contribution of this work is the application of the proposed pose estimation method to real mission data. Although the synthetic dataset presented in Section 3 was generated using a rendering tool capable of producing high-fidelity imagery, an inevitable domain gap remains between synthetic and real observations. The real data used in this study consist of image sequences of the JAXA H-2A upper stage acquired during a fly-around inspection conducted by the demonstration satellite ADRAS-J, developed by Astroscale Japan [2]. No additional data have been released, including camera parameters or ground-truth pose annotations, which prevents a fully quantitative evaluation. To address this limitation, a camera calibration procedure was conducted to estimate the camera parameters from a set of 2D–3D correspondences between image and CAD model keypoints. The calibration algorithm employs an iterative Levenberg–Marquardt optimization to minimize the reprojection error defined in Equation (2),

$$argmin_{K,R,\mathbf{t}} \sum_{i=1}^{n} ||\mathbf{p}^i - \pi(\mathbf{p}_w^i, K, R, \mathbf{t})||^2 \qquad (2)$$

where $\pi(\mathbf{p}_w^i, K, R, \mathbf{t})$ denotes the reprojection of the 3D point $\mathbf{p}_w^i$ onto the image plane using the intrinsic matrix K, rotation matrix R and translation vector $\mathbf{t}$. A total of 19 images were used to extract point correspondences for the calibration process, resulting in a Root Mean Square (RMS) reprojection error of 1.295 pixels. The best-performing model configuration identified in Section 4 was subsequently used to estimate the pose of the H-2A upper stage using the calibrated camera parameters. Qualitative examples of pose predictions are presented in Figure 6. The results indicate that rotations around the main axis (visualized in blue) are not consistent across consecutive frames. This behavior was expected due to the high degree of rotational symmetry of the target, which exhibits a predominantly cylindrical geometry and lacks distinctive features capable of disambiguating rotations around its symmetry axis.

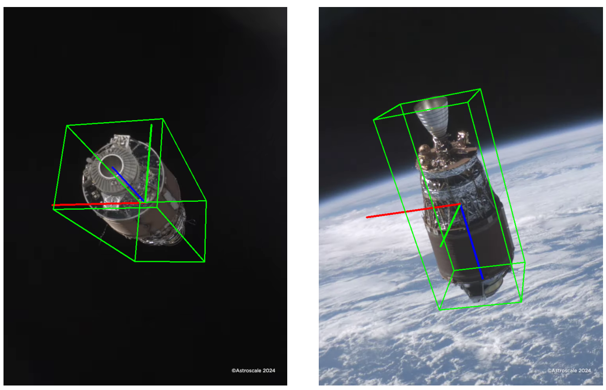

Figure 6: Pose estimation visualizations of the H-2A upper stage.

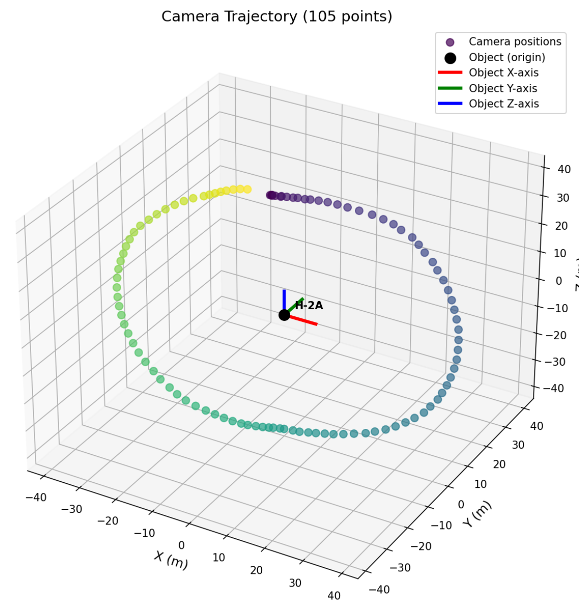

Figure 7: Smoothed camera trajectory visualization.

Although the predicted poses cannot be quantitatively evaluated due to the absence of ground-truth annotations, they can be exploited to estimate the trajectory of the servicer spacecraft relative to the target, thereby providing an additional qualitative form of validation. This analysis relies on several assumptions: the target is not spinning, the rotation around the main symmetry axis is fixed to mitigate the rotational ambiguity induced by axial symmetry, and the camera frame rate is constant, which is required for the subsequent trajectory smoothing step. The reconstruction is performed by inverting the predicted target poses to obtain the camera poses expressed in the target reference frame. The resulting camera positions are then used to compute the camera-to-target distances, yielding an average distance of 44.736 meters. This value is consistent with the publicly reported distance of approximately 50 meters, considering the uncertainty introduced by the estimated camera parameters. To obtain a smoother trajectory estimate, a window-based Moving Average filter was applied to the 105 reconstructed trajectory points. Figure 7 illustrates the smoothed trajectory obtained using a window size of 15. The reconstructed motion appears physically plausible and consistent with the expected fly-around maneuver.

6. Conclusions

This thesis investigates the application of a state-of-the-art terrestrial unseen object pose estimation method to spacecraft imagery, targeting relative navigation and in-orbit proximity operations with non-cooperative spacecraft. The proposed approach relies on a deep learning–based framework that does not require object-specific training, thereby enabling generalization to previously unseen targets. Due to the absence of a multi-target spacecraft dataset specifically designed for pose estimation, a high-fidelity synthetic dataset was generated using multiple 3D spacecraft models sourced from the ESA public repository. The experimental results and subsequent analysis provide several important insights into the challenges of unseen spacecraft pose estimation in orbit, highlighting in particular the critical role of textured rendering in resolving rotational ambiguities and the difficulties posed by highly symmetric targets. The application to real mission data further demonstrated the robustness of the proposed approach and enabled the reconstruction of a physically plausible elliptical fly-around trajectory consistent with publicly available mission information. Overall, this work shows that generalizable deep learning pose estimation models can be effectively transferred to the space domain, enabling relative navigation with non-cooperative spacecraft without requiring target-specific training.

References

[1] Michele Bechini, Michèle Lavagna, and Paolo Lunghi. Dataset generation and validation for spacecraft pose estimation via monocular images processing. *Acta Astronautica*, 2023.

[2] Japanese Aerospace Exploration Agency (JAXA). CRD2 Phase I / ADRAS-J Update: Fly-Around Observation Images of Space Debris Released, 2024. `https://global.jaxa.jp/press/2024/07/20240730-1_e.html`. Visited: 22-02-2026.

[3] Yann Labbé, Lucas Manuelli, et al. Megapose: 6d pose estimation of novel objects via render compare. `https://arxiv.org/abs/2212.06870`, 2022.

[4] Tae Ha Park, Marcus Martens, et al. Speed+: Next-generation dataset for spacecraft pose estimation across domain gap. In *2022 IEEE Aerospace Conference (AERO)*. IEEE, 2022.

[5] Leo Pauly, Wassim Rharbaoui, et al. A survey on deep learning-based monocular spacecraft pose estimation: Current state, limitations and prospects. *Acta Astronautica*, 212:339–360, November 2023.