International Conference on Robotics, Machine Learning and Artificial Intelligence (ICRMLAI), 2024

“EFFECTIVE LANDMARK REGRESSION USING ATTENTION BASED-HRNET FOR SATELLITE POSE ESTIMATION”

By Abraham Paul, Anoop GL, Ganesh Kumar R,
Aryan Singh, Sri Aditya Deevi, Ravikumar Lagisetty

PRESENTED BY,
ABRAHAM PAUL

WHAT IS SATELLITE POSE ESTIMATION?

- **Satellite Pose Estimation** focuses on determining the **position, orientation or rotational state** of a satellite in space using images or other sensor data.

- It is crucial for space missions, especially when performing tasks like **satellite servicing, docking, debris removal or inspection**.

MOTIVATION

- The ability to accurately estimate the pose of satellites is a critical requirement in modern space missions. With the growing demand for autonomous satellite operations, such as servicing, docking and debris removal precise orientation and position determination have become increasingly important.

- Traditional methods of pose estimation often struggle with accuracy in dynamic and visually complex environments, leading to the need for more sophisticated approaches.

Autonomous Satellite Operations -> **The increasing demand for autonomous space missions, including satellite servicing and docking requires precise and reliable pose estimation systems.**

Space Debris Removal -> **As space debris accumulates, accurate satellite pose estimation becomes critical for debris removal missions, ensuring successful navigation in cluttered environments.**

Improvement Over Traditional Methods -> Existing pose estimation techniques often lack accuracy, especially in dynamic conditions. Advanced computer vision models can improve precision and robustness.

Enhancement of Spacecraft Maneuverability -> Accurate pose estimation helps in optimizing the satellite's navigation and control, contributing to the overall efficiency of space missions.

OBJECTIVES OF MY RESEARCH

- Develop a ***vision-based satellite pose estimation framework leveraging deep learning techniques*** with a primary objective of accurately determining the orientation of satellites relative to a reference frame

- Implement an ***effective landmark regression using attention based HRNet*** to localize and track specific landmarks or features on satellite and determine the pose

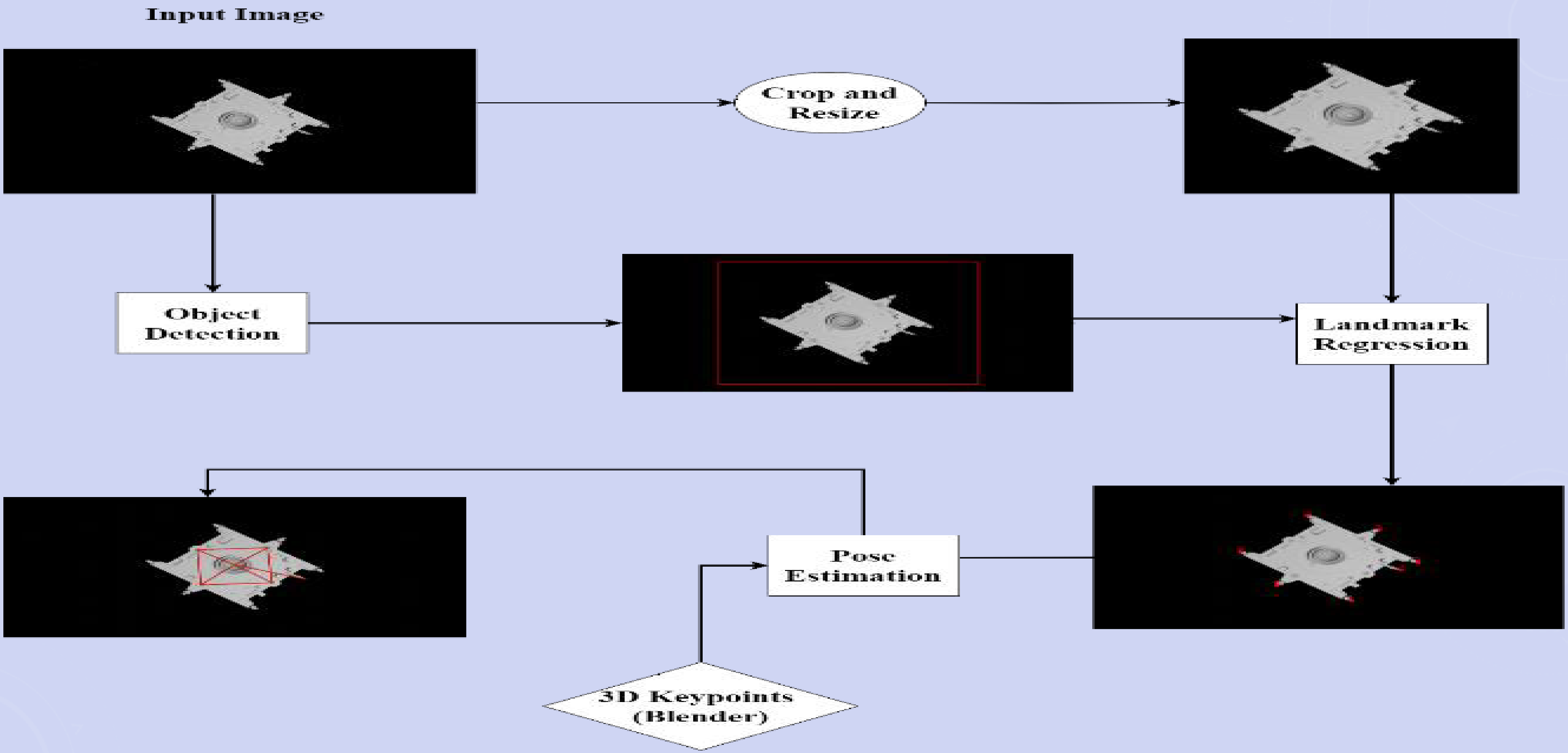
GAP ANALYSIS

- Limited Accuracy in Traditional Methods
- Insufficient Utilization of Deep Learning Models
- Challenges in Real-Time Pose Estimation
- Lack of Comprehensive Datasets for Space Scenarios
- Integration with Spacecraft Systems

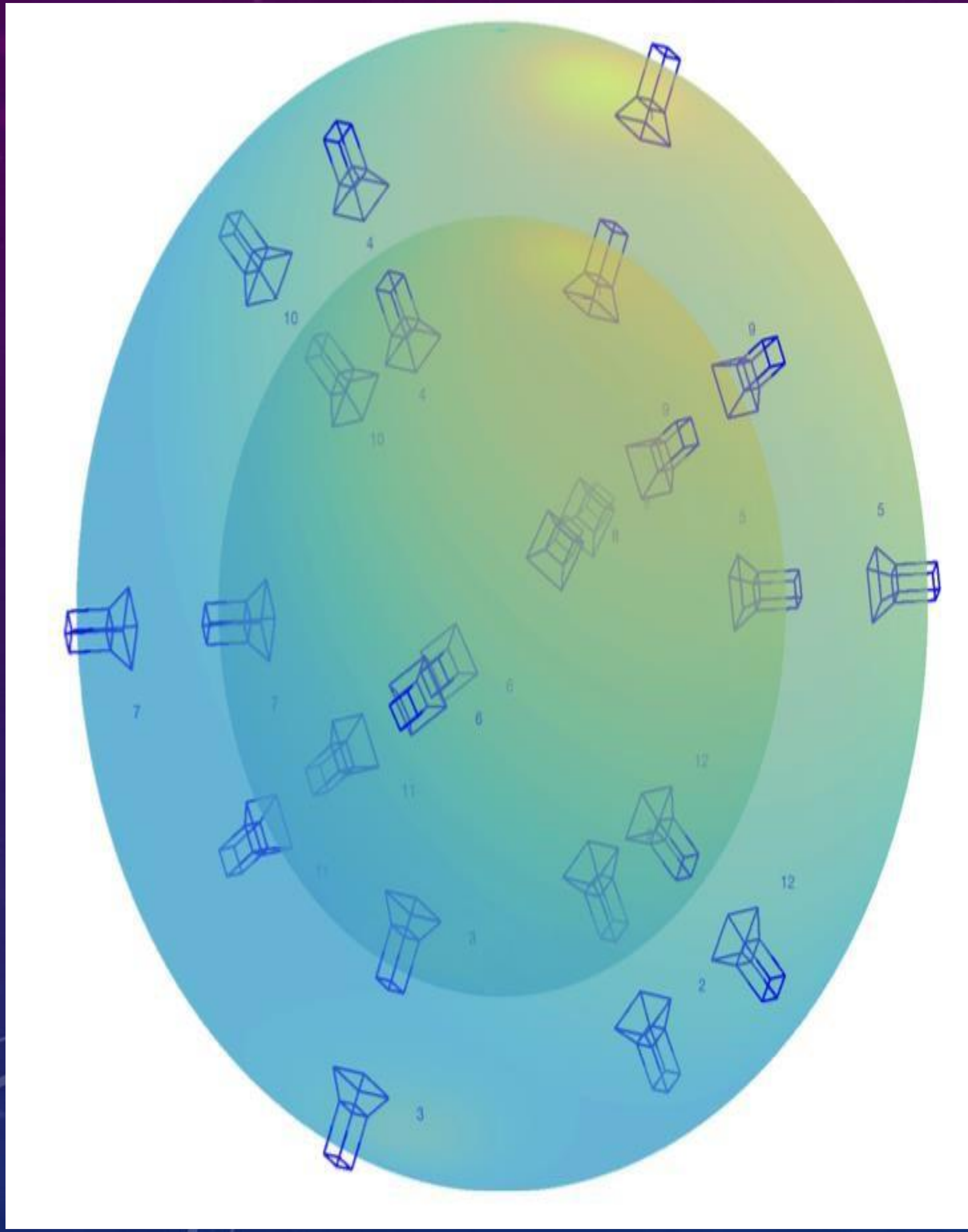
PROBLEM STATEMENT

This research aims to address these challenges by developing *a vision-based satellite pose estimation system that leverages Faster-RCNN for object detection, High-Resolution Networks (HRNet) for keypoint detection and the Perspective-n-Point (PnP) algorithm for pose calculation* providing a robust and accurate solution suitable for real-time applications in satellite operations.

METHODOLOGY



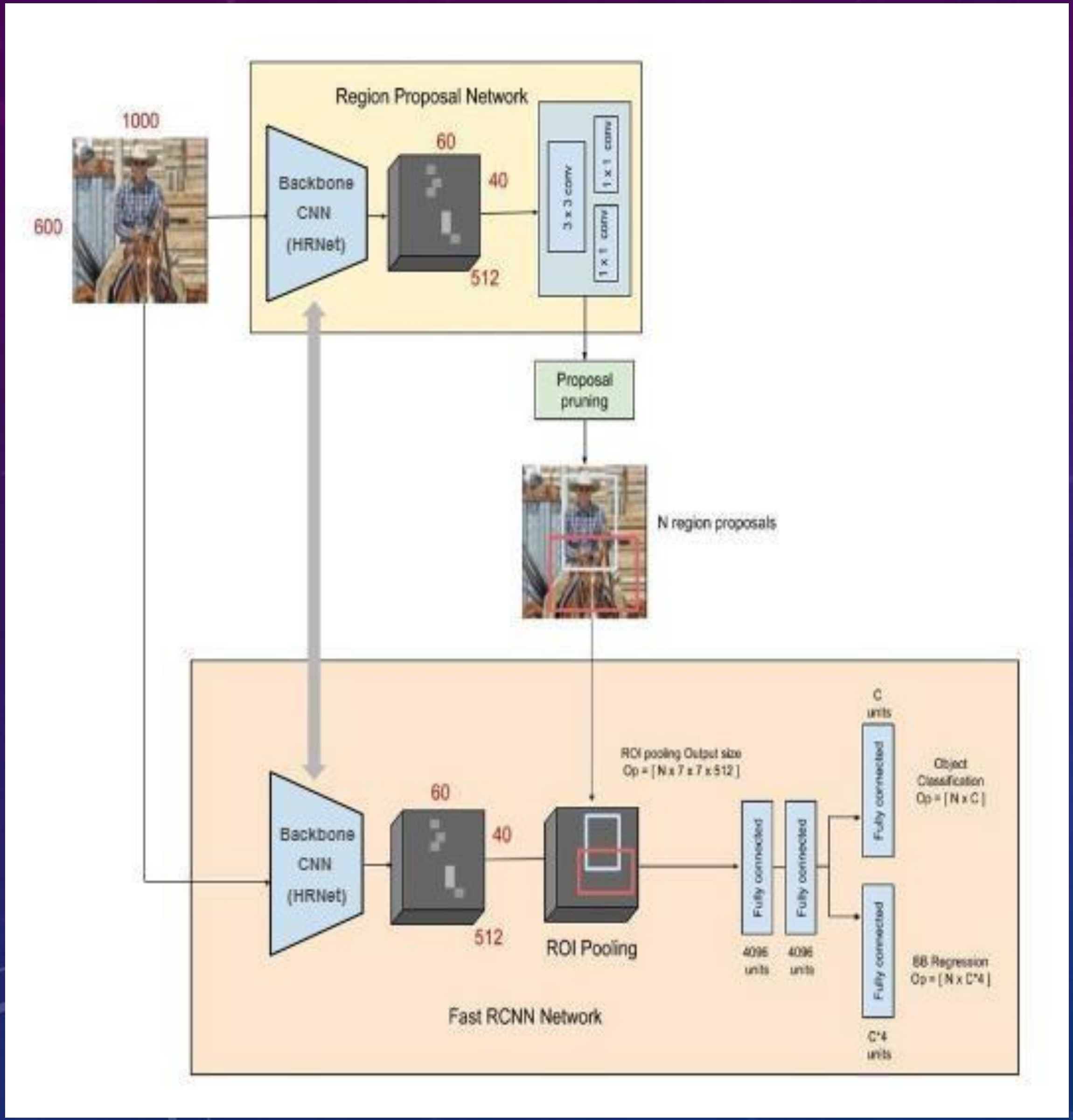
DATASET GENERATION



The dataset of **GSAT-12R satellite model** which is diverse was generated using **Blender** tool by incorporating an element of systematic randomness to ensure proper training of the model:

- The range between the **camera** and the **satellite** is between **3 meters and 19 meters**
- The **camera positions were restricted** to only a part of the sphere where **$0 \leq \phi \leq 80$**
- **Thompson sampling technique** was used to ensure camera positions have good sphere coverage
- The **camera positions were perturbed** to ensure the object is not always at center of image which can cause redundancy in the dataset
- The camera was also **rotated along boresight axis** to ensure the dataset is diverse and all the images where captured with different poses

FASTER –RCNN FOR OBJECT DETECTION

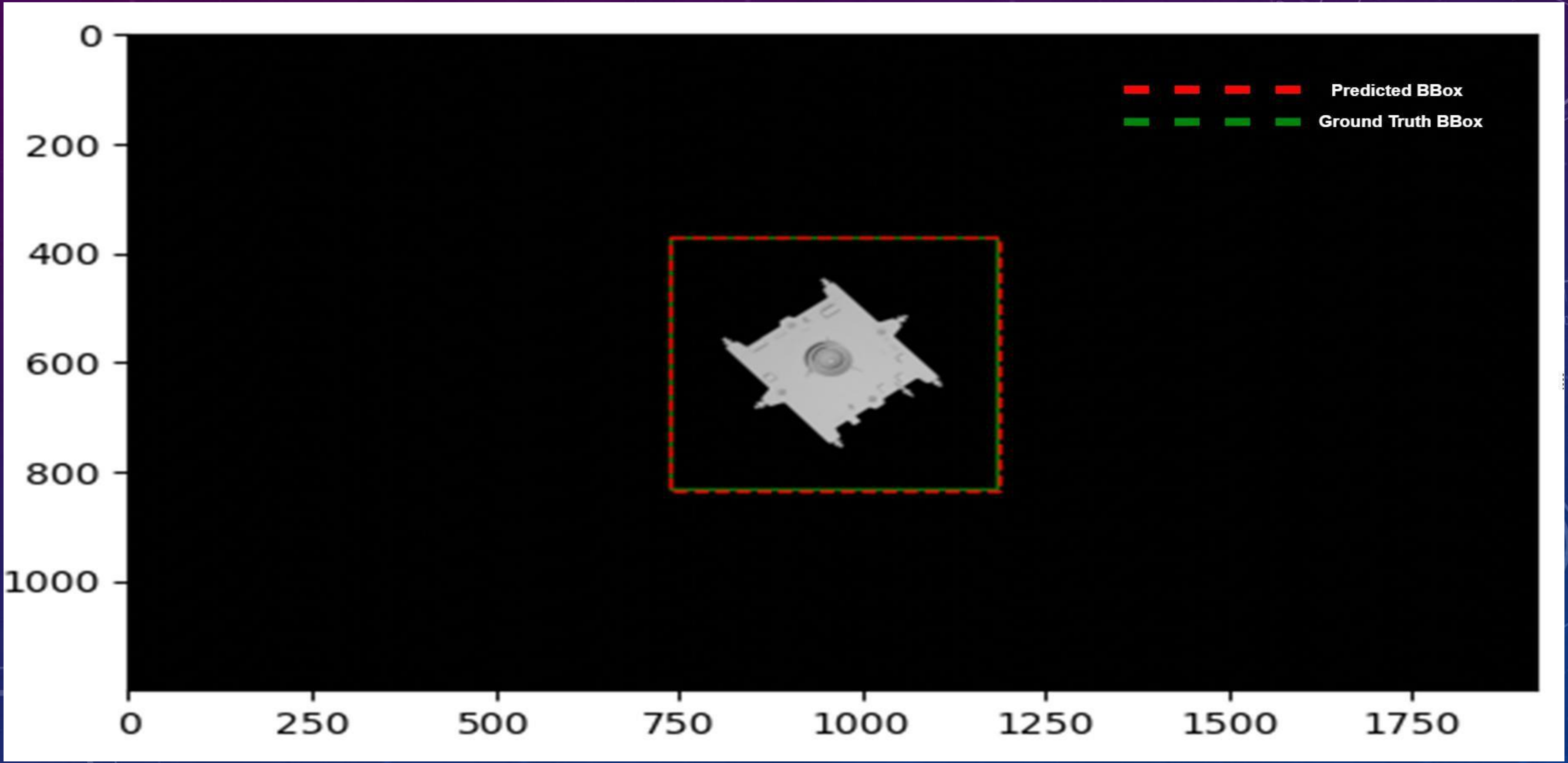


Faster R-CNN with an HRNet backbone is a powerful model for object detection, combining **precise region proposal generation** with detailed object classification and bounding box refinement.

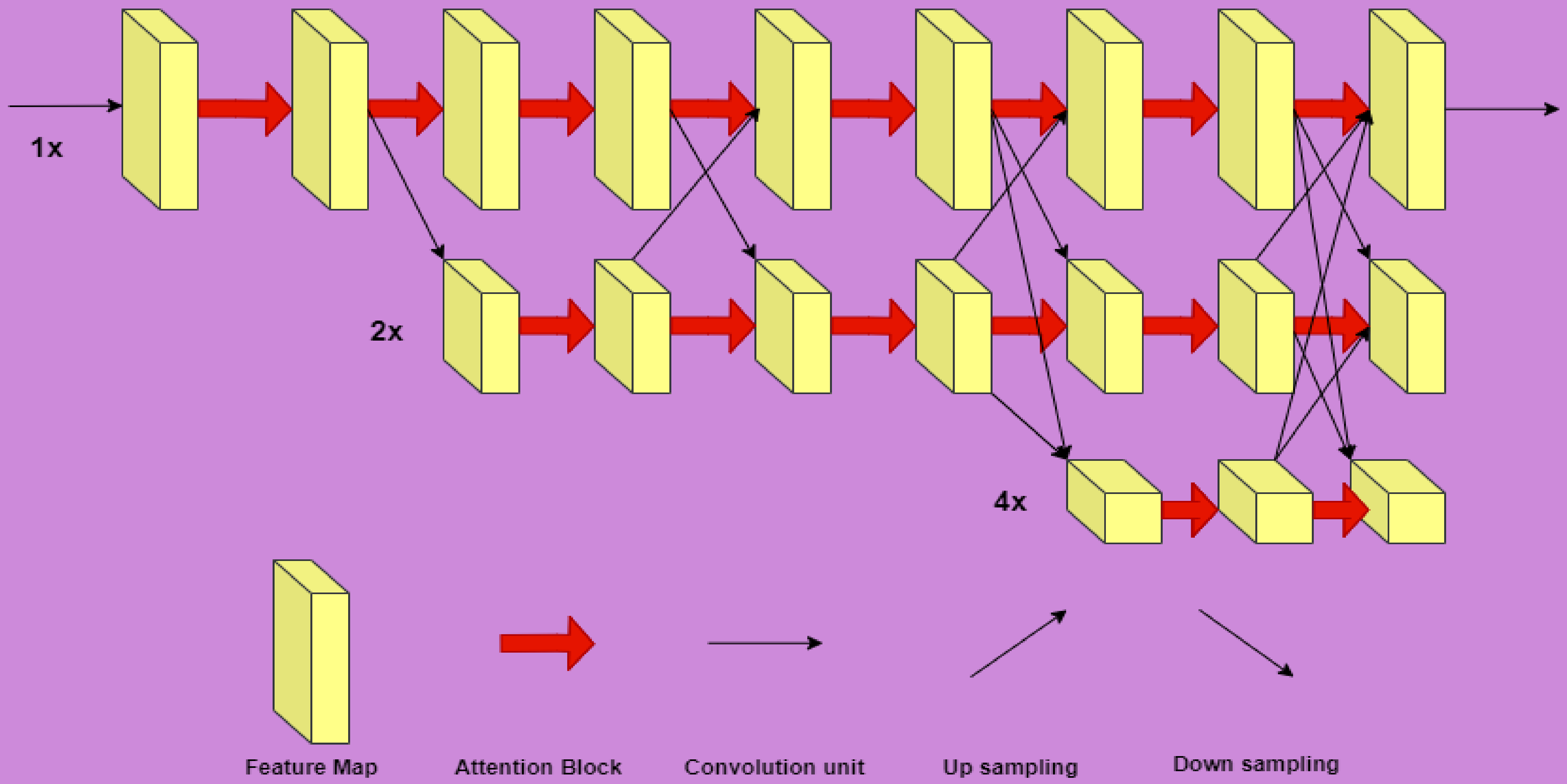
HRNet as the backbone, **maintains high-resolution feature maps**, capturing fine spatial details that standard networks may miss. This high-resolution and multi-scale fusion enhances Faster R-CNN's ability to **localize objects accurately**.

In satellite pose estimation, this combination enables the model to better identify complex details. Ultimately, HRNet's detailed feature representation **boosts the overall accuracy and spatial fidelity** of Faster R-CNN's object detection capabilities.

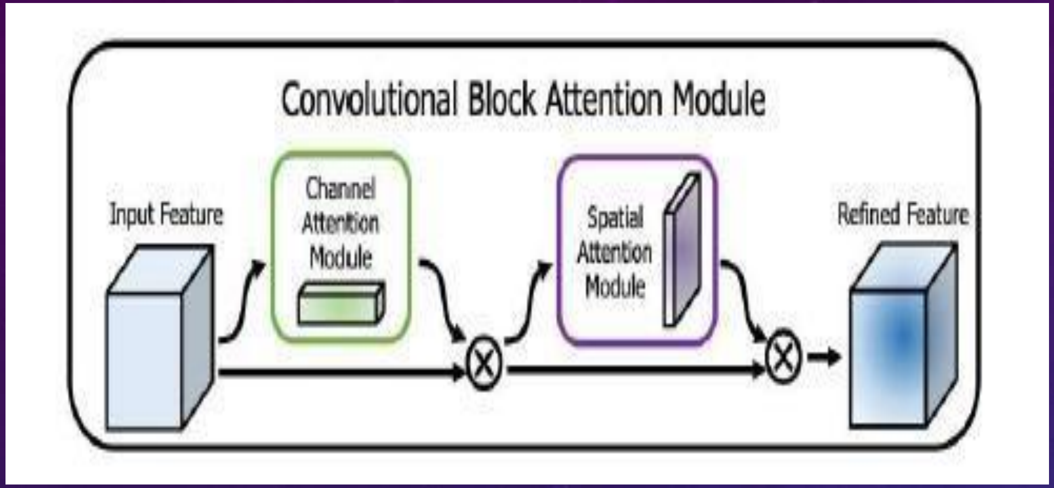
RESULT OF OBJECT DETECTION



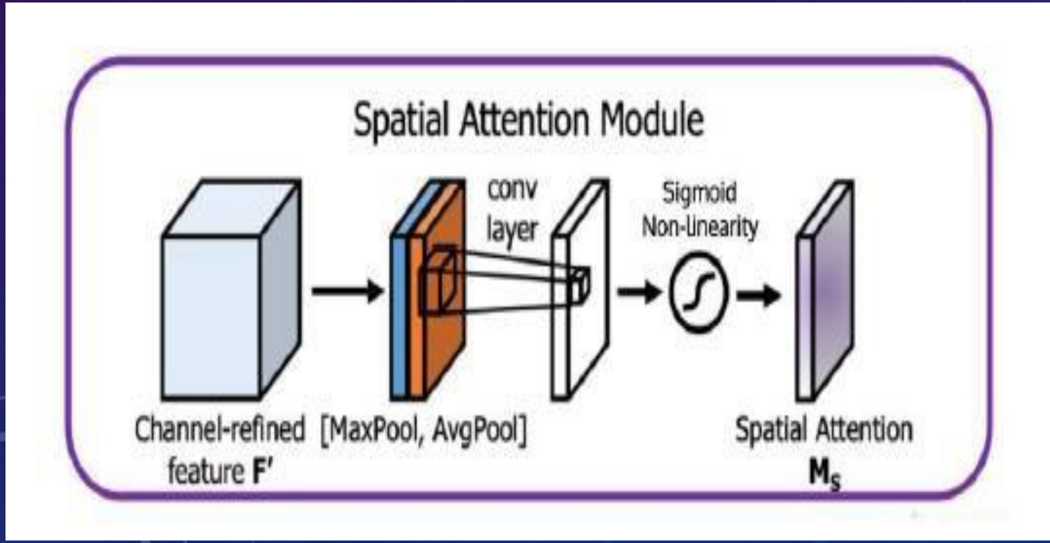
AHRNET ARCHITECTURE FOR LANDMARK REGRESSION



ATTENTION BLOCKS IN AHRNET



Channel Attention: channel attention sub-module is to **improve the feature maps** by cross-channel interaction



Spatial Attention: it is helps the model to basically **decide "where" to focus in a feature map**

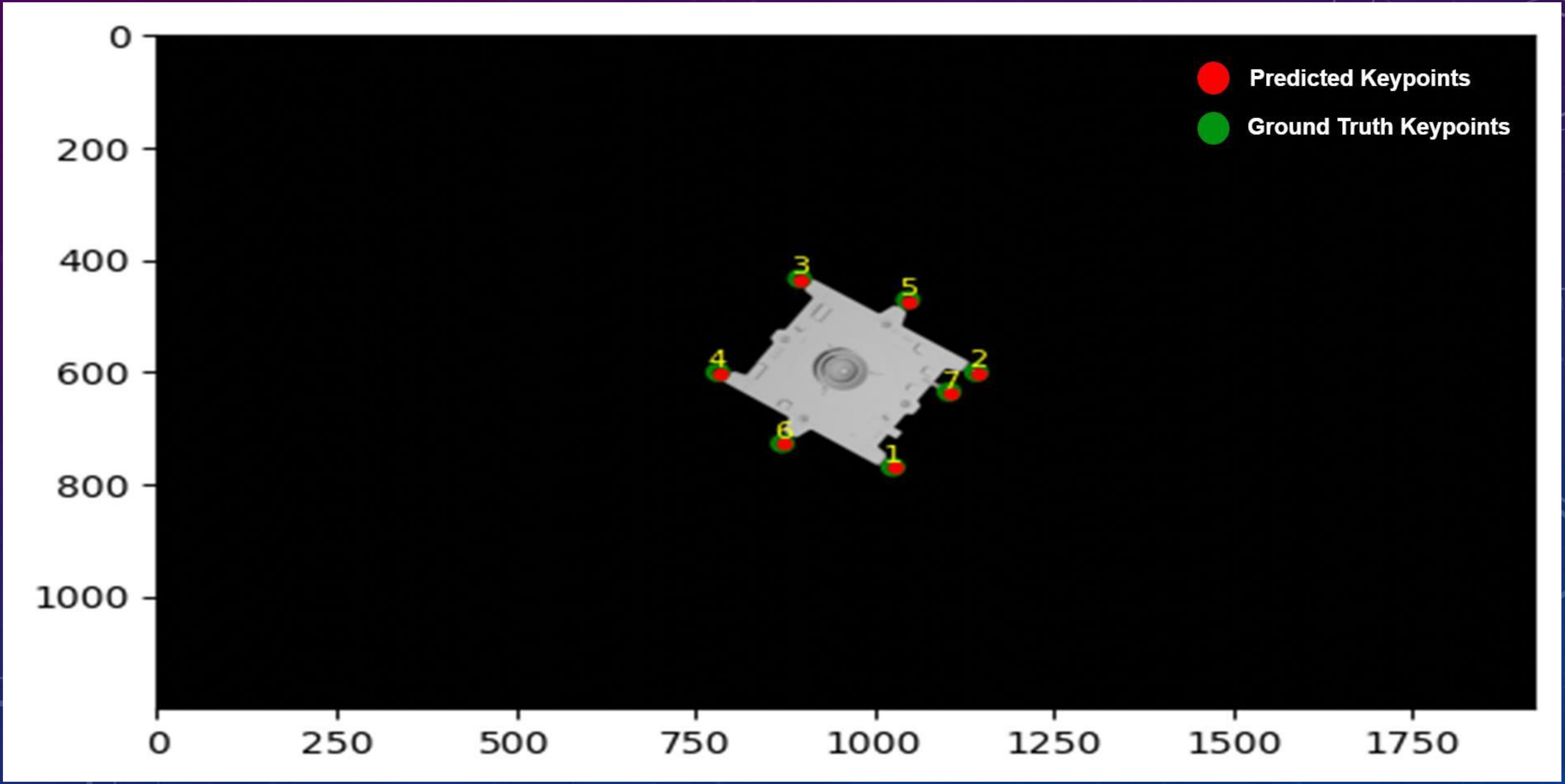
WHY AHRNET FOR LANDMARK REGRESSION?

- **Attention-Based HRNet (AHRNet)** enhances HRNet by adding attention mechanisms to focus on crucial areas in an image for **precise landmark regression**.

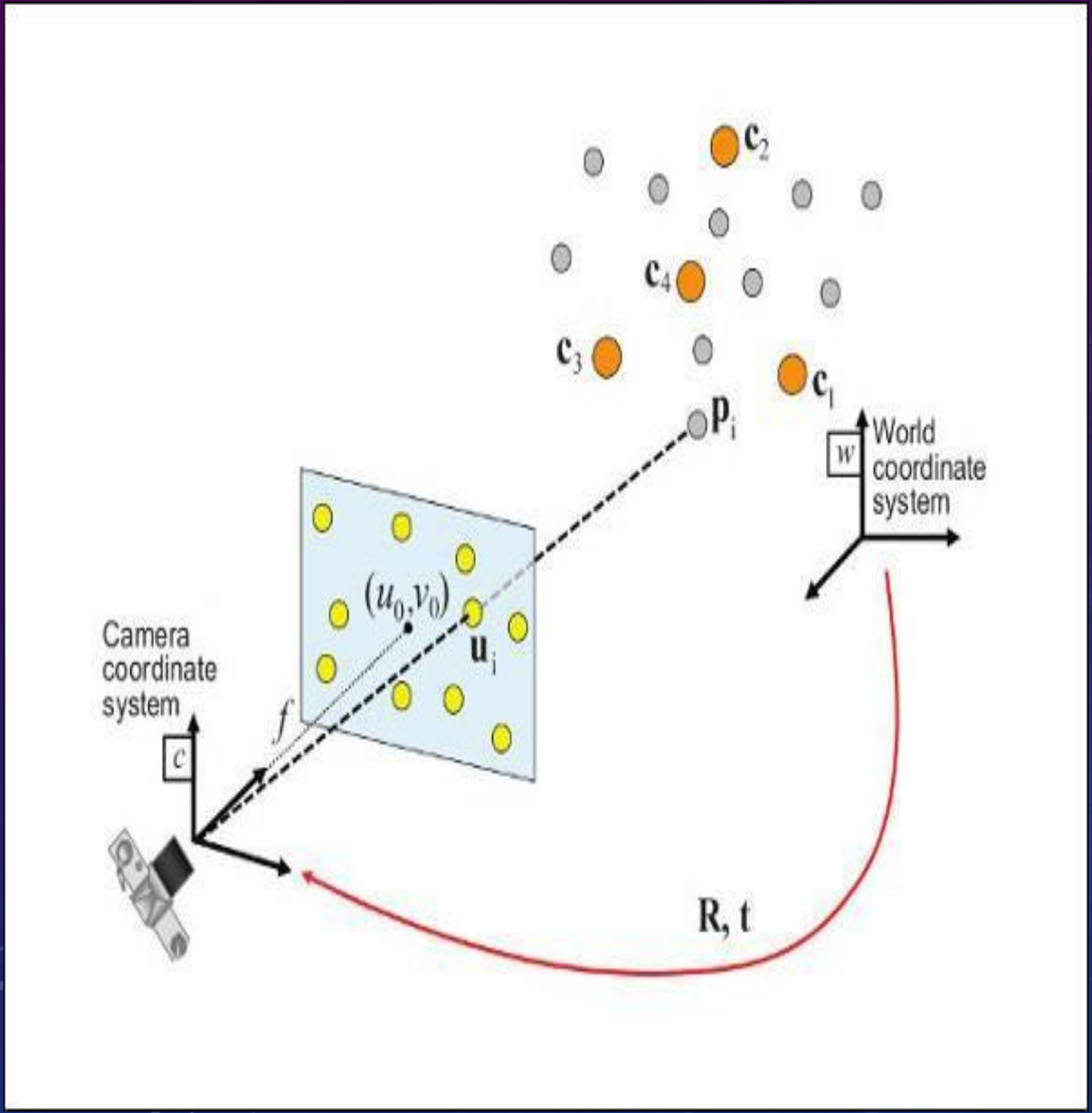
- It maintains **high-resolution feature maps and uses attention to highlight important regions, reducing background noise and irrelevant details**. Multi-scale feature fusion with attention allows AHRNet to **capture details across various scales, improving landmark localization accuracy.**

- This approach is particularly effective in complex images, achieving more accurate keypoint detection. In satellite pose estimation, **AHRNet's precision in detecting landmarks enables reliable 2D-3D mapping, essential for accurate pose calculation**.

THE KEYPOINTS OF THE SATELLITE MODEL



POSE ESTIMATION USING PNP ALGORITHM



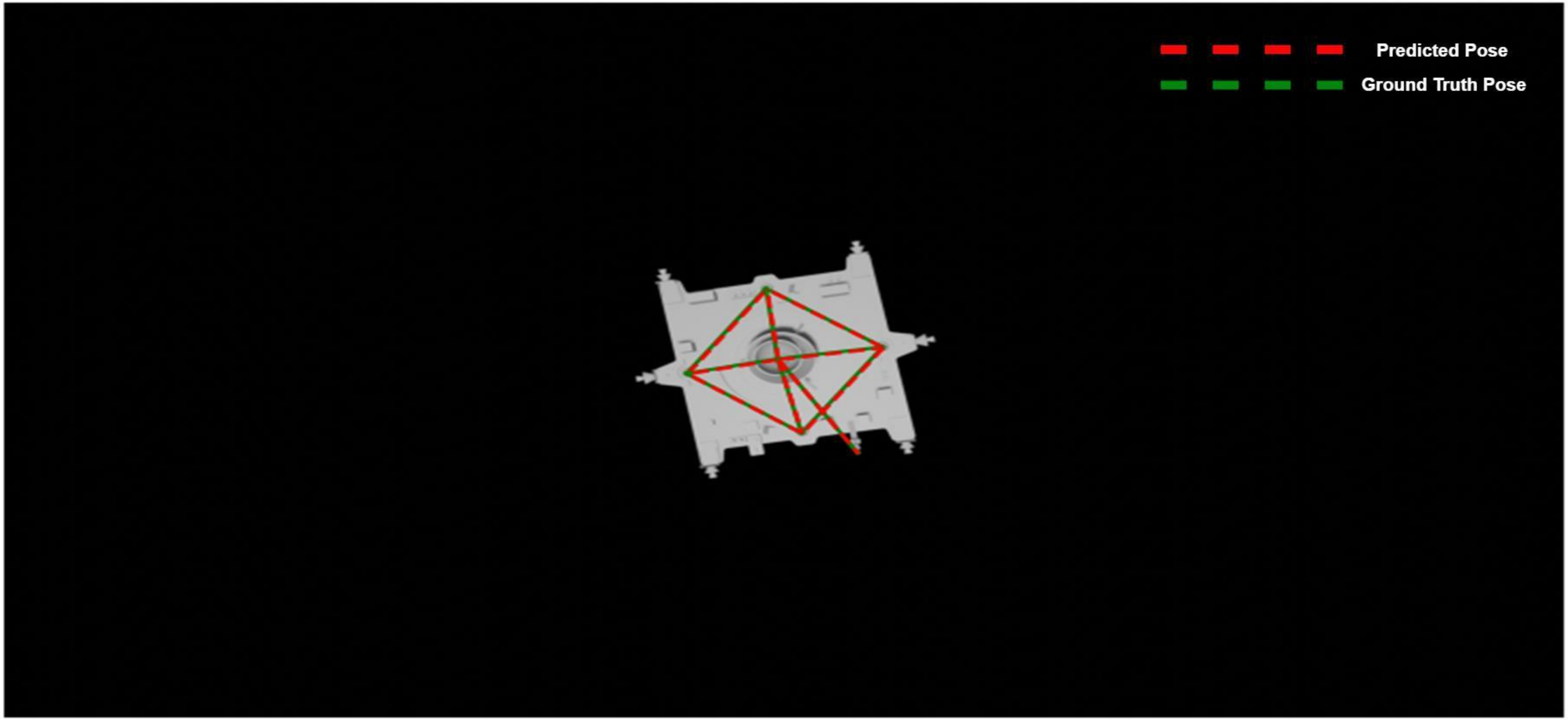
Pose estimation using the PnP algorithm involves calculating the 3D position and orientation of an object relative to a camera.

It uses **two main inputs**: known 3D points on the object (satellite) and their corresponding 2D projections in the image. Camera intrinsic parameters, like focal length help relate the 3D points to the 2D plane accurately.

The goal of PnP is to determine rotation and translation vectors that align the 3D points with their 2D counterparts. By minimizing reprojection error (difference between observed and projected points), the algorithm refines these vectors.

Optimization methods like **Levenberg-Marquardt or RANSAC handle noise or outliers** during this process. The output provides the object's pose, essential for applications needing precise spatial orientation such as robotics and satellite tracking.

RESULTS OF POSE ESTIMATION



ROTATION AND TRANSLATION ERROR CALCULATION

We define E_R and E_T as follows:

> $E_R = | eul^* - eul |$, where

E_R is the rotation error,

eul^* is the ground truth rotation value,

eul is the estimated rotation value

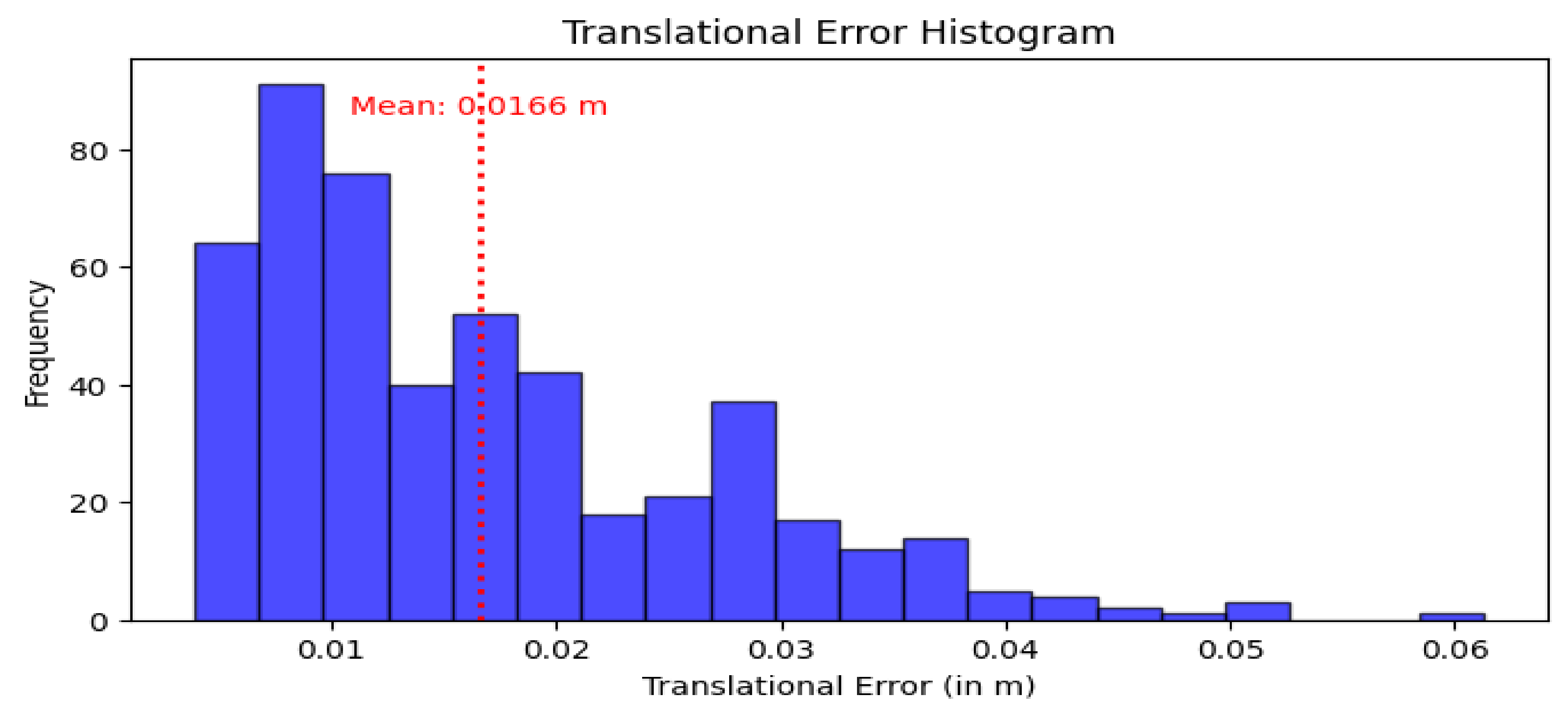
> $E_T = || tr^* - tr ||^2$, where

E_T is the translation error,

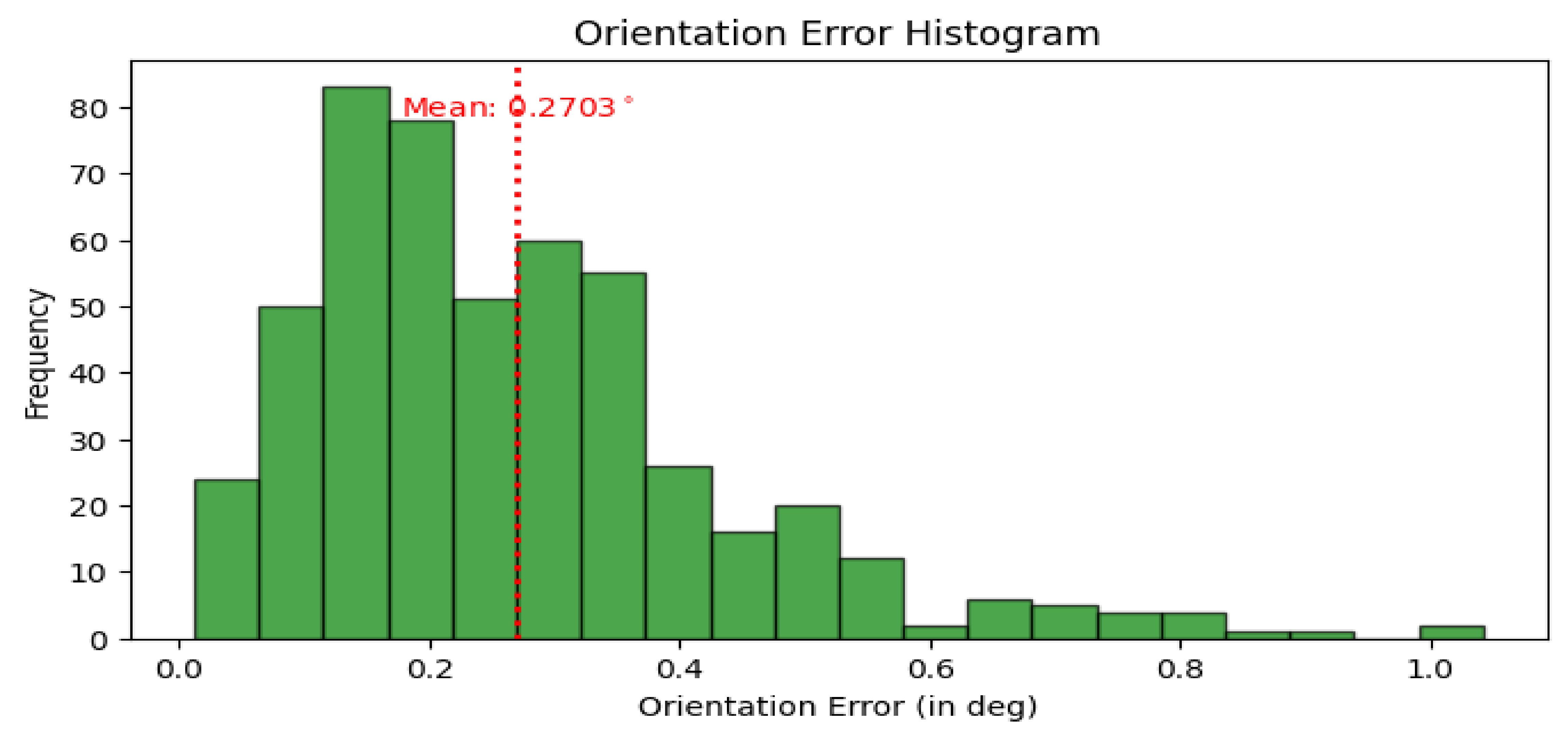
tr^* is the ground truth translation value,

tr is the estimated translation value

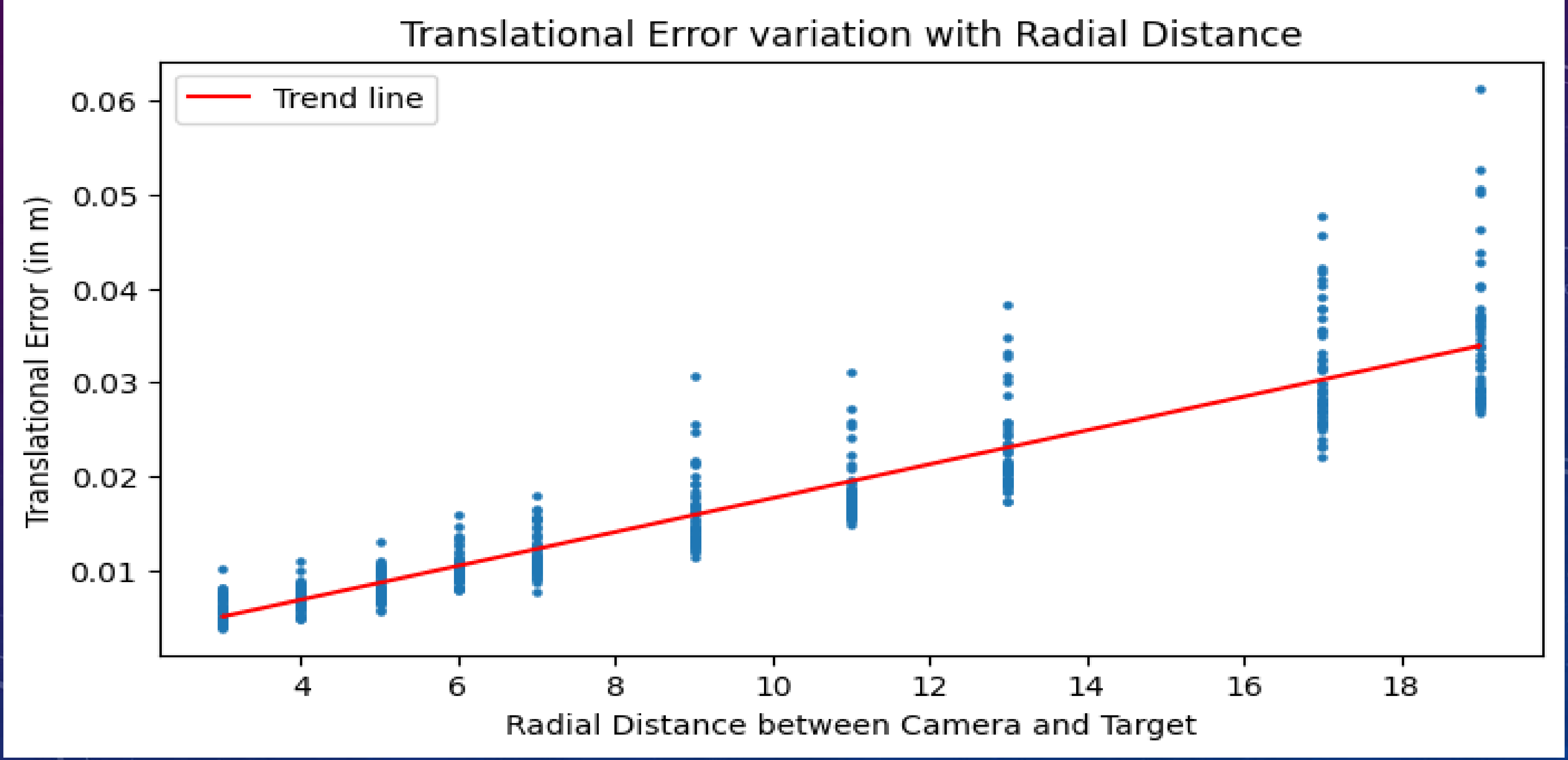
RESULTS - TRANSLATION ERROR VS FREQUENCY



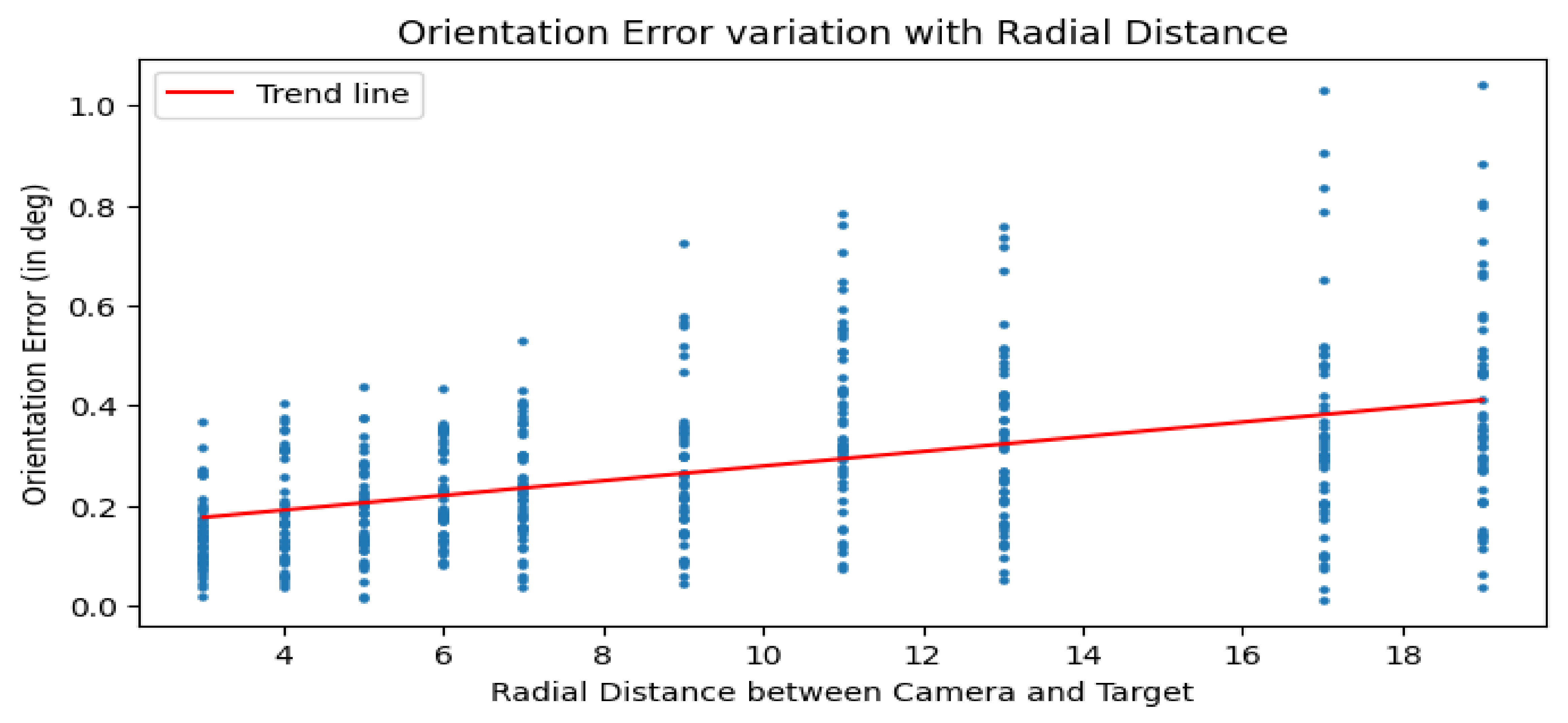
RESULTS - ORIENTATION ERROR VS FREQUENCY



RESULTS - TRANSLATION ERROR VS RADIAL DISTANCE



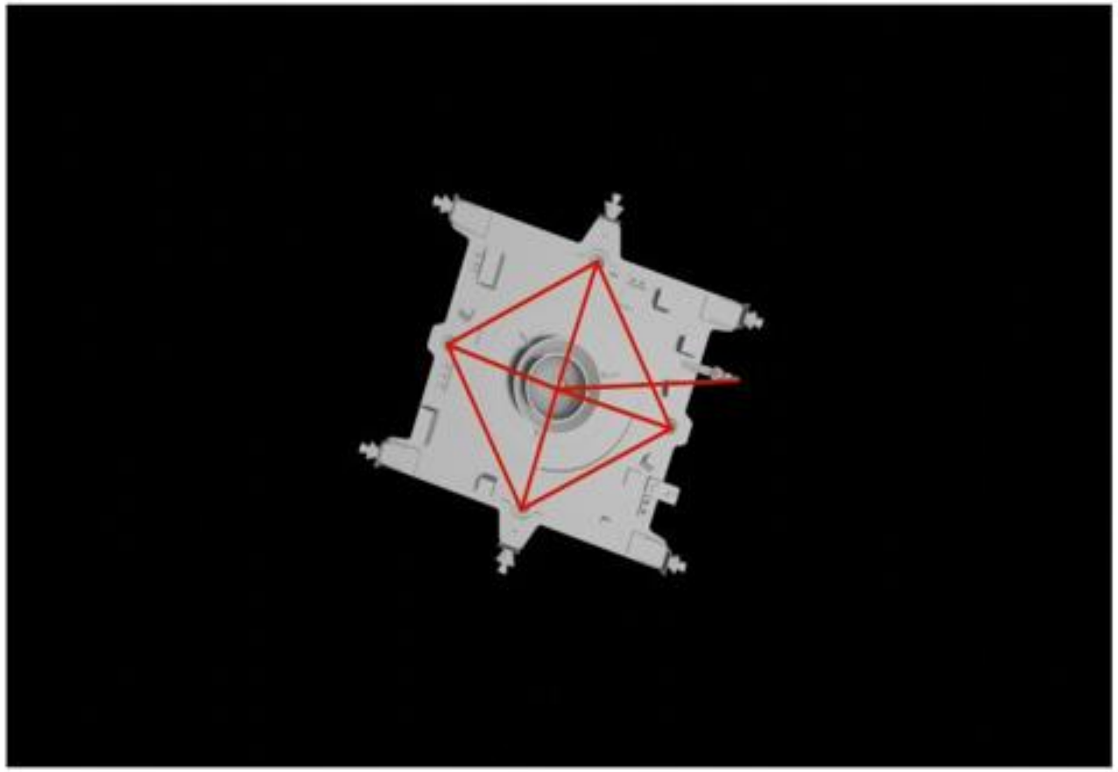
RESULTS - ORIENTATION ERROR VS RADIAL DISTANCE



Verification of the pose estimation results

Green Wired Frame = Ground Truth

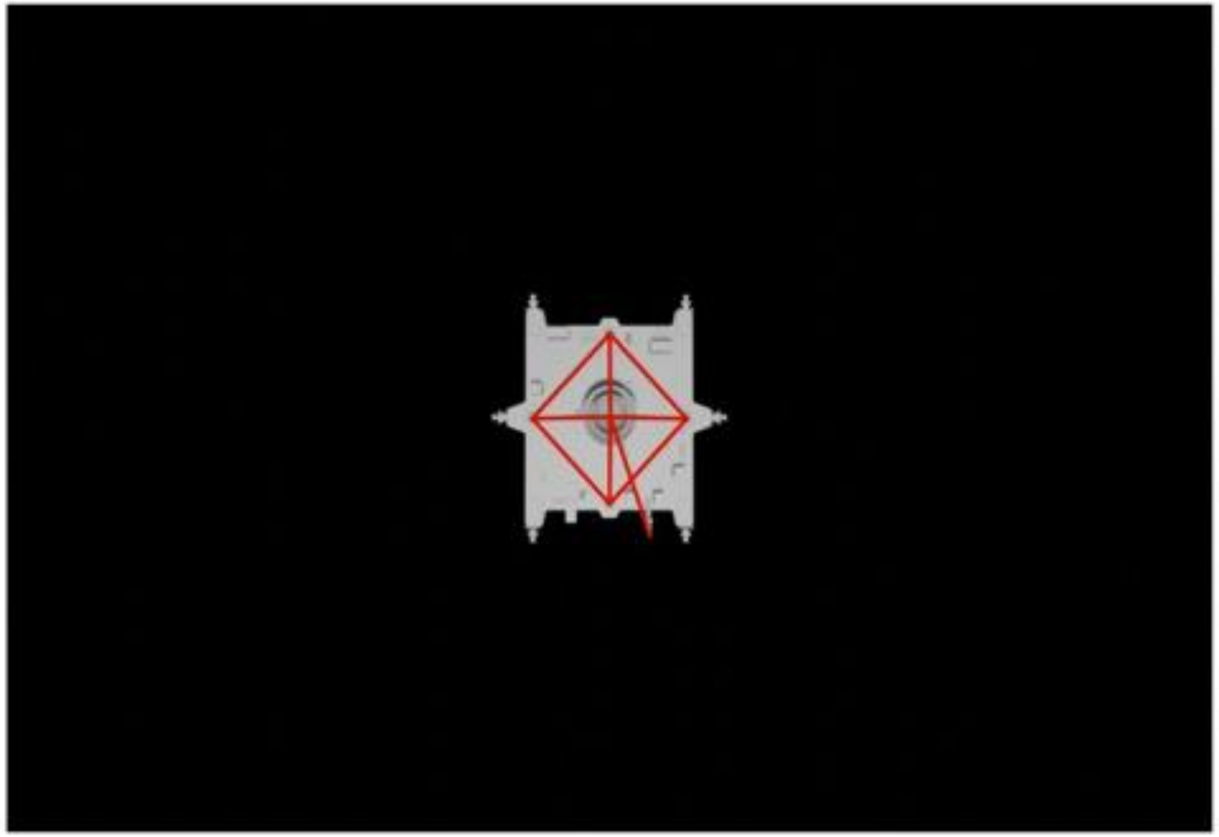
Red Wired Frame = Predicted



Translation Error: 0.002605

Rotational Error: 0.093684 degrees

Distance from camera to satellite: 3.000000238418588



Translation Error: 0.006566

Rotational Error: 0.265582 degrees

Distance from camera to satellite: 5.0

RESULTS

Sl	Error Metric	Value
1	Mean Translational Error (E_T) *[in m]*	0.016617009651114525
2	Mean Rotation Error (E_R) *[in deg]*	0.27031458666224356
3	Mean Absolute Translational Error ($\|t_{gt} - t_{pred}\|$) *[in m]*	0.00988876 (X) 0.00972417 (Y) 0.00770925 (Z)
4	Mean Absolute Orientation Error ($\|eul_{gt} - eul_{pred}\|$) *[in deg]*	0.37910969 (X) 0.38968987 (Y)

Sl	Reprojection Error	Value (in pixels)
1	Reprojection Error for E1	2.0339158276185025
2	Reprojection Error for E4	1.9305897301039263
3	Reprojection Error for W1	2.07580426463015 4
4	Reprojection Error for W4	1.5296732370876553
5	Reprojection Error for ST	1.77574739516093 1
6	Reprojection Error for NT	1.8886139873071606
7	Reprojection Error for NFAOut	1.6476353685377223
	Overall Reprojection Error	**1.474331614309867 3**

FUTURE SCOPE

- In the future, we plan to implement **satellite pose tracking** for tracking the satellite for each and every frame. While this concept is well-established for objects on Earth, its application in space is still in its early development stages and offers significant opportunities for advancement.
- We also plan to use **dense correspondence methods** which will further improve the reprojection error, translation and rotation error for estimation of the pose of the satellite.
- Focusing on **enhancing the model's domain adaptability** to handle various real-world conditions such as **unpredictable sunlight, reflections** and **diverse surface textures** is also our future work. Improving adaptability in these environments would result in more precise pose estimation in actual space missions.

CONCLUSION

- In a nutshell, we have **successfully implemented satellite pose estimation using a new methodology** which resulted in less error and improved accuracy during pose estimation.

- Using AHRNet, we were able to **perform landmark regression and detect the key points in the satellite with a higher accuracy** rate. Our methodology resulted in better accuracy with a low error rate after training the model for just 10 epochs.

REFERENCES

[1] *J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, "Deep High- Resolution Representation Learning for Visual Recognition" arXiv preprint arXiv:1908.07919, 2020.*

[2] *B. Chen, J. Cao, A. Parra and T. -J. Chin, "Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement" 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South), 2019, pp. 2816-2824, doi: 10.1109/IC- CVW.2019.00343. technologies" Proc. 8th Int. Symp. Artif. Intell., Robot. Autom. Space, 2005.*

3 *H. Zhou, T. Zhang, and J. Jayender, "Re-weighting and 1-Point RANSAC-Based PnP Solution to Handle Outliers" arXiv preprint arXiv:2007.08577, 2020.*

4 *L. P. Cassinis, R. Fonod, and E. Gill, "Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft" Prog. Aerosp. Sci., 2019.*

5 *S. D'Amico, M. Benn, and J. L. Jørgensen, "Pose estimation of an uncooperative spacecraft from actual space imagery" Int. J. Space Sci. Eng., vol. 2, no. 2, pp. 171-189, 2014.*

6 *S. Sharma and S. D'Amico, "Pose estimation for non-cooperative rendezvous using neural networks" in AAS/AIAA Astrodynamics Specialist Conference, 2019.*

REFERENCES (CONT.)

7 S. Sharma, J. Ventura, and S. D'Amico, "Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous" J. Spacecr. Rockets, vol. 55, no. 6, pp. 1414-1429, 2018.

8 A. Kendall and R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization" in Proc. IEEE Int. Conf. Robot. Autom. (ICRA), 2016.

9 A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning" in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2017.

10 A. Kendall, M. Grimes, and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization" in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015.

11 V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An accurate o(n) solution to the PnP problem" Int. J. Comput. Vis., vol. 81, no. 2, pp. 155-166, 2009.

12 I. Melekhov, J. Ylioinas, J. Kannala, and E. Rahtu, "Image-based localization using hourglass networks" in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017.

13 M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns" IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 3, pp. 448-461, 2009.

14 S. Peng, Y. Liu, Q. Huang, X. Zhou, and H. Bao, "Pvnet: Pixel-wise voting network for 6DOF pose estimation" in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

REFERENCES (CONT.)

15 T. Sattler, Q. Zhou, M. Pollefeys, and L. Leal-Taixe, “Understanding the limitations of CNN-based absolute camera pose regression” in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019.

16 K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation”, 2019.

17 S. Sharma, J. Ventura, and S. D’Amico, “Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous” J. Spacecr. Rockets, vol. 55, no. 6, pp. 1414-1429, 2018.

18 K. Sun, B. Xiao, D. Liu, and J. Wang, “Deep high-resolution representation learning for human pose estimation” in Proc. IEEE Conf. Comput. Vis. Pattern Recog. (CVPR), 2019, pp. 5693-5703.

19 T. H. Park, M. Ma¨rttens, G. Le´cuyer, D. Izzo, and S. D’Amico, “SPEED+: Next Generation Dataset for Spacecraft Pose Estimation across Domain Gap”, arXiv preprint arXiv:2110.03101, 2021.

20 E. Shreyas, M. H. Sheth, and Mohana, “3D Object Detection and Tracking Methods using Deep Learning for Computer Vision Applications” in Proc. IEEE RTEICT, 2021, doi: 10.1109/RTEICT52294.2021.9573964.

21 S. Sharma, C. Beierle and S. D'Amico, “Pose estimation for non-cooperative spacecraft rendezvous using convolutional neural networks," 2018 IEEE Aerospace Conference, Big Sky, MT, USA, 2018, pp. 1-12, doi: 10.1109/AERO.2018.8396425.

22 Deevi, Sri Aditya, and Deepak Mishra. “Expeditious object pose estimation for autonomous robotic grasping." International Conference on Computer Vision and Image Processing. Cham: Springer Nature Switzerland, 2022.

THANK YOU