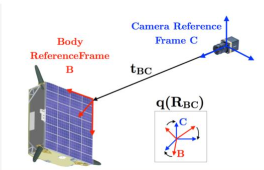
# Robust Learning-Based Pose Estimation of Noncooperative Spacecraft

Tae Ha (Jeff) Park
tpark94@stanford.edu

Video presentation available at https://youtu.be/fUSmC93MkVk

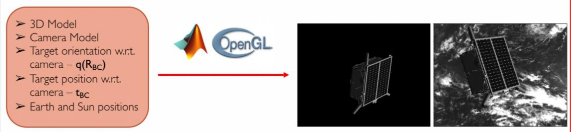
## POSE ESTIMATION

Use **a single monocular camera** from a servicer spacecraft to determine a **pose** of the **non-cooperative** target relative to the servicer's camera




## DATASET

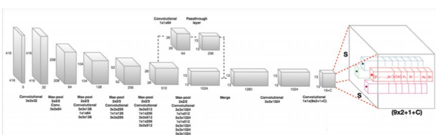
- Spacecraft PosE Estimation Dataset (SPEED) [1] – synthetically generate photorealistic spacecraft images




- 21 images from PRISMA mission with labeled pose (PRISMA-21)

## CNN ARCHITECTURE

The project uses a bounding box regression technique using the network by Tekin et al. [2] based on YOLOv2 [3]. This network outputs nine 2D points corresponding to the box centroid and its eight corners. Then, these corners can be used in conjunction with known 3D model to solve Perspective-n-Points (PnP) problem in order to extract relative attitude and position.



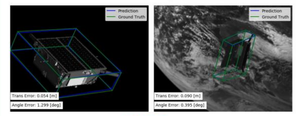
## RANDOMIZED TEXTURE

Ideally, the network is trained using synthetic images due to extreme difficulty of acquiring real spacecraft images with annotated pose. However, the network trained on synthetic dataset tend to lack robustness when applied to real images. Among many differences between two image sources, the biggest difference is the **texture** of the target spacecraft. In order to force the network invariant to varying spacecraft surface textures, the style augmentation technique [4] is used to apply **random neural style transfer** to the spacecraft images. This effectively randomizes the spacecraft texture, leading to improved robustness when applied to real images.

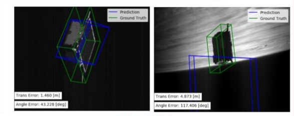
## RESULTS & DISCUSSIONS

1. Trained on SPEED synthetic images *without* texture randomization

| Metrics | SPEED | PRISMA-21 |
|---|---|---|
| MPE [pix] | 3.615 | 113.5 |
| Mean $E_T$ [m] | [ 0.009, 0.008, 0.188 ] | [ 0.366, 0.138, 1.715 ] |
| Median $E_T$ [m] | [ 0.006, 0.005, 0.101 ] | [ 0.153, 0.074, 1.665 ] |
| Mean $E_R$ [deg] | 2.202 | 61.396 |
| Median $E_R$ [deg] | 1.818 | 22.395 |



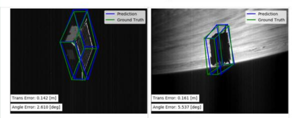
SPEED



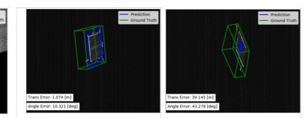
PRISMA-21

2. Trained on SPEED synthetic images *with* texture randomization

| Metrics | SPEED | PRISMA-21 |
|---|---|---|
| MPE [pix] | 6.291 | 25.678 |
| Mean $E_T$ [m] | [ 0.024, 0.023, 0.655 ] | [ 0.106, 0.160, 3.972 ] |
| Median $E_T$ [m] | [ 0.011, 0.009, 0.216 ] | [ 0.028, 0.030, 0.704 ] |
| Mean $E_R$ [deg] | 4.907 | 11.334 |
| Median $E_R$ [deg] | 3.330 | 6.273 |



Improvement!



Failure Cases

We see that training with texture-randomized synthetic images, the network performs pose estimation on real images with improved accuracy *without having been trained on them*. Since textures cannot be leveraged, the network is focusing on other features that are common to both data sources – **shape**. This explains the failure cases in which the spacecraft occluded due to adverse illumination from Sun or Earth.

[1] Sharma S., D'Amico S. "Pose Estimation for Non-Cooperative Rendezvous Using Neural Networks," 2019 AAS/AIAA Astrodynamics Specialist Conference, Ka'anapali, Maui, HI, January 13-17 (2019).
[2] Tekin B., Sinha S.N., and Fua P., "Real-Time Seamless Single Shot 6D Object Pose Prediction", CVPR 2018.
[3] Redmon J., Farhadi A. "YOLO9000: Better, Faster, Stronger," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017): 6517-6525.
[4] Jackson P.T.G., Abarghouei A.A., Bonner S., Breckon T.P., Obara B., "Style Augmentation: Data Augmentation via Style Randomization," CoRR abs/1809.05375 (2018)