CVLAB UNSEEN SPACECRAFT POSE ESTIMATION

Jérémy CHAVEROT

École polytechnique fédérale de Lausanne

SUPERVISOR: Dr. Mathieu SALZMANN

ADVISORS: Dr. Andrew PRICE, PhD. Chen ZHAO

19th January 2024

Jérémy Chaverot

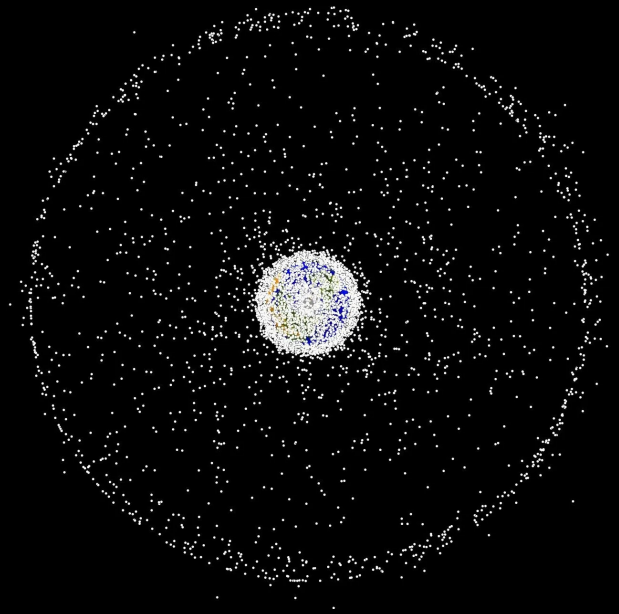
Objective

- Unseen 6DoF Pose Competition with ESA Advanced Concept Team
- Given an image, estimate the relative 3D translation and 3D rotation of the spacecraft w.r.t. the camera
- Synthetic images generated by the Spacecraft Dataset team
  - Four spacecrafts models
  - Earth rendered background
  - Access to models, masks, segmentations, camera settings and ground truth poses

Jérémy Chaverot

Applications

- Remediation of space debris in Low Earth Orbit (LEO)
- Space monitoring
- Planetary defense technology (e.g. NASA's DART Mission)



Jérémy Chaverot

OUTLINE

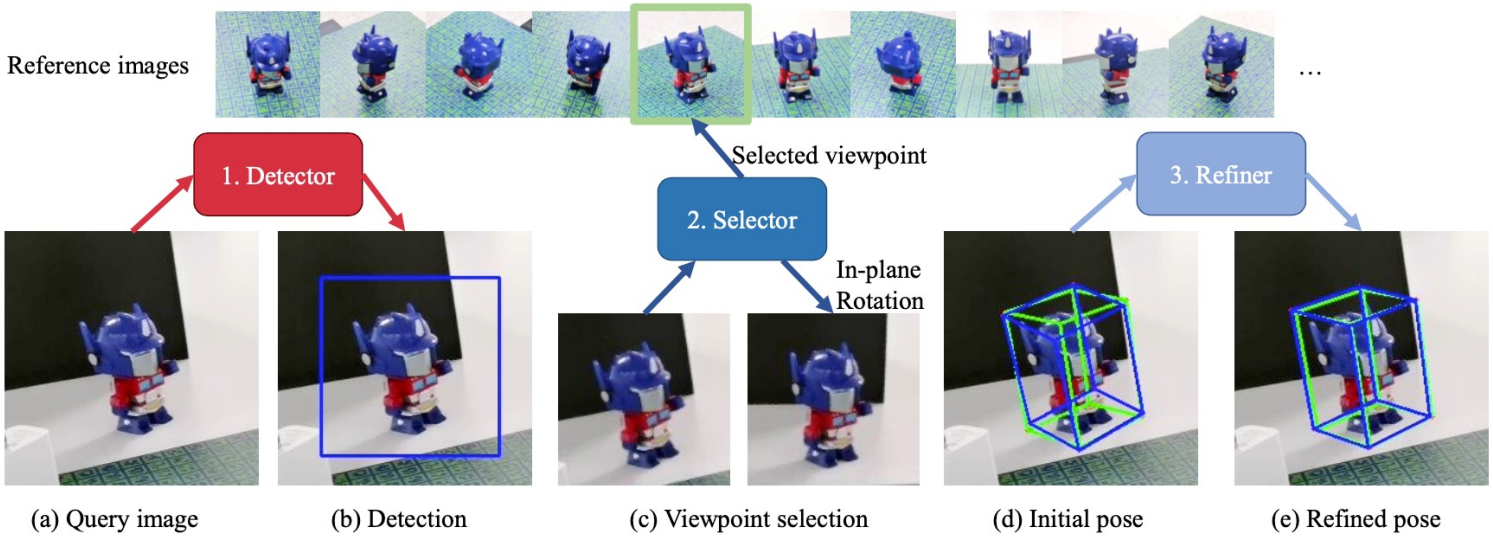
Gen6D

Implementation

Evaluation

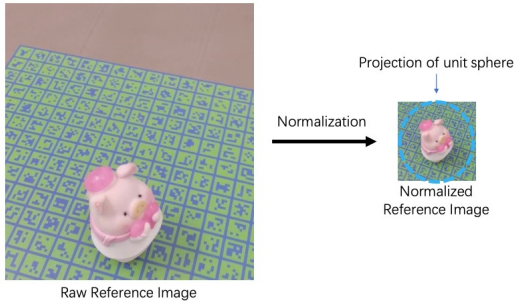
Amelioration of the results

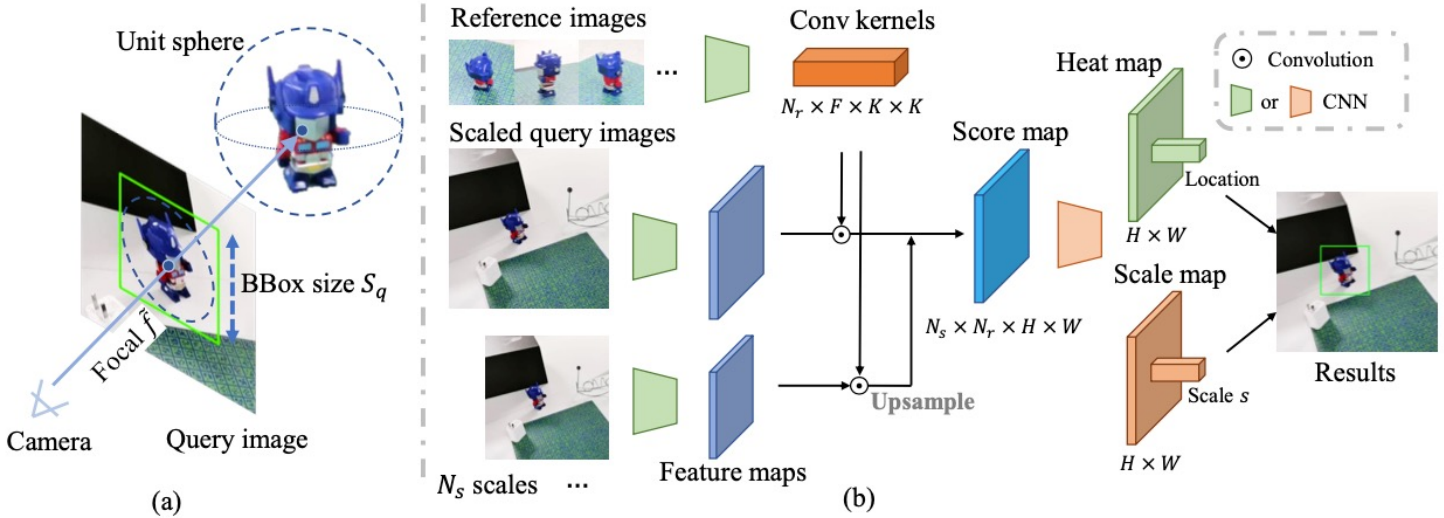
Gen6D, how it works



Y. Liu, Y. Wen, S. Peng, C. Lin, X. Long, T. Komura, and W. Wang. Gen6D: Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images. 2023. arXiv: 2204.10776

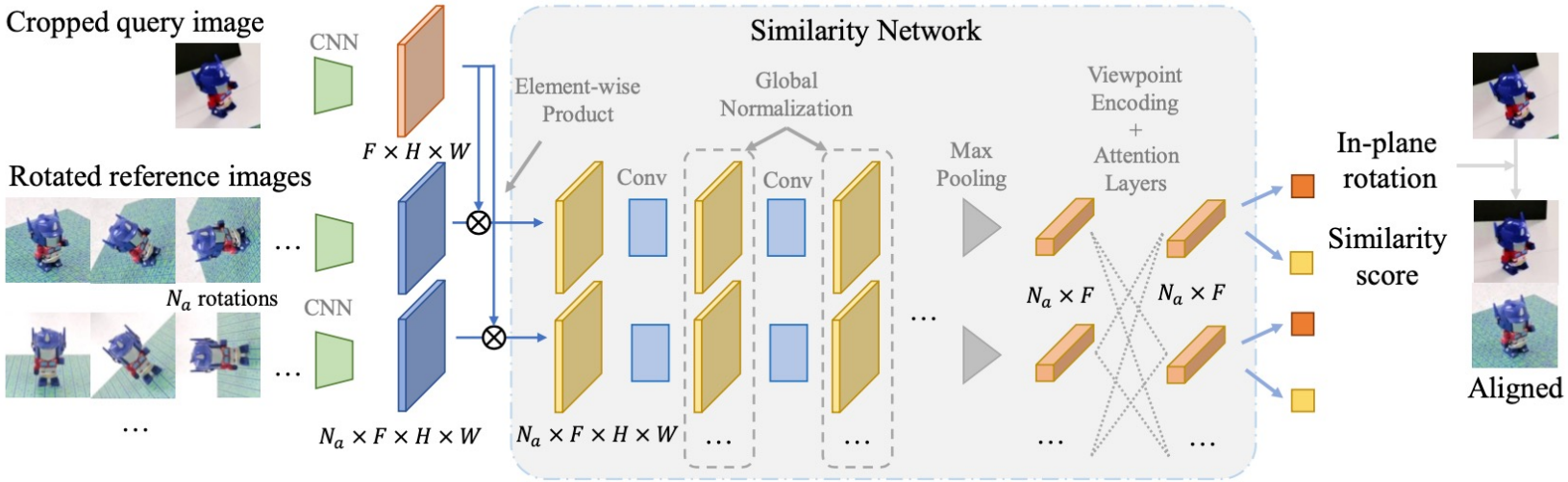
Gen6D, the detector





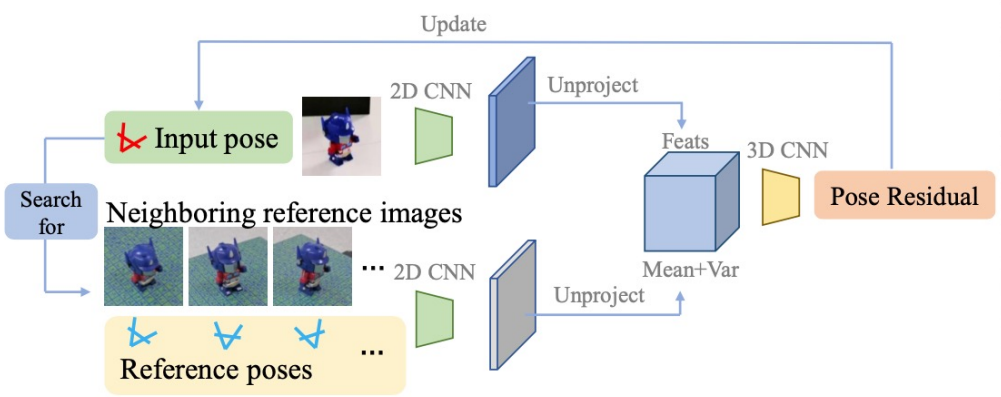
Y. Liu, Y. Wen, S. Peng, C. Lin, X. Long, T. Komura, and W. Wang. Gen6D: Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images. 2023. arXiv: 2204.10776

Gen6D, the viewpoint selector

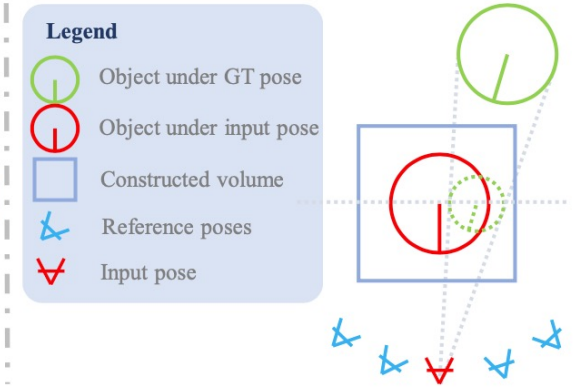


Y. Liu, Y. Wen, S. Peng, C. Lin, X. Long, T. Komura, and W. Wang. Gen6D: Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images. 2023. arXiv: 2204.10776

Gen6D, the pose refiner



(a) Architecture of the pose refiner



(b) Similarity approximation

Y. Liu, Y. Wen, S. Peng, C. Lin, X. Long, T. Komura, and W. Wang. Gen6D: Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images. 2023. arXiv: 2204.10776

Gen6D, strengths and weaknesses

Table 1: Summary of Gen6D

Pros	Cons
Generalizability	Limited by Reference Image Quality
Model-Free	Everyday Life Objects Training Data
Simple Input Requirements	Difficulty with Symmetric Objects
Robustness to Background Clutter	Dependence on Initial Detection and Selection
Effective in Diverse Environments	Potential Challenges with Severe Occlusions
Competitive Performance	Computationally Intensive

Y. Liu, Y. Wen, S. Peng, C. Lin, X. Long, T. Komura, and W. Wang. Gen6D: Generalizable Model-Free 6-DoF Object Pose Estimation from RGB Images. 2023. arXiv: 2204.10776

OUTLINE

Gen6D

Implementation

Evaluation

Amelioration of the results

SCITAS Izar, the environment

- Equipped with two NVIDIA V100 PCIe 32 GB GPUs
- Ideal to run efficiently a Machine Learning task by anyone and anywhere
- Faced some technical challenges:
  - Setting up the virtual environment
  - Installing the necessary dependencies
  - Composing the bash execution script
  - Fundamentally, learning the correct way to utilize the server
- Special recognition to Emily Bourne from the EPFL HPC team

Jérémy Chaverot

Steps to implement Gen6D

- Redefinition of Gen6D's data loader
- Convert the quaternions into their corresponding rotation matrix
- Invert the masks
- Resize the query images

OUTLINE

Gen6D

Implementation

Evaluation

Amelioration of the results

Jérémy Chaverot

The evaluation metrics

The estimated pose: $\hat{\mathbf{P}} = (\hat{\mathbf{R}}, \hat{\mathbf{T}})$

The ground truth pose: $\bar{\mathbf{P}} = (\bar{\mathbf{R}}, \bar{\mathbf{T}})$

$$e_{\text{ADD}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}, \mathcal{M}) = \underset{\mathbf{x} \in \mathcal{M}}{\text{avg}} \left\| \bar{\mathbf{P}}\mathbf{x}^{\star} - \hat{\mathbf{P}}\mathbf{x}^{\star} \right\|_2$$

$$= \underset{\mathbf{x} \in \mathcal{M}}{\text{avg}} \left\| (\bar{\mathbf{R}}\mathbf{x} + \bar{\mathbf{T}}) - (\hat{\mathbf{R}}\mathbf{x} + \hat{\mathbf{T}}) \right\|_2$$

$$e_{\text{ADD-S}}(\hat{\mathbf{P}}, \bar{\mathbf{P}}, \mathcal{M}) = \underset{\mathbf{x}_1 \in \mathcal{M}}{\text{avg}} \underset{\mathbf{x}_2 \in \mathcal{M}}{\min} \left\| \bar{\mathbf{P}}\mathbf{x}_1^{\star} - \hat{\mathbf{P}}\mathbf{x}_2^{\star} \right\|_2$$

$$= \underset{\mathbf{x}_1 \in \mathcal{M}}{\text{avg}} \underset{\mathbf{x}_2 \in \mathcal{M}}{\min} \left\| (\bar{\mathbf{R}}\mathbf{x}_1 + \bar{\mathbf{T}}) - (\hat{\mathbf{R}}\mathbf{x}_2 + \hat{\mathbf{T}}) \right\|_2$$

S. Hinterstoisser, V. Lepetit, S. Ilic, S. Holzer, G. Bradski, K. Konolige, and N. Navab. “Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes.” In: Computer Vision – ACCV 2012.

Jérémy Chaverot

Too large query images

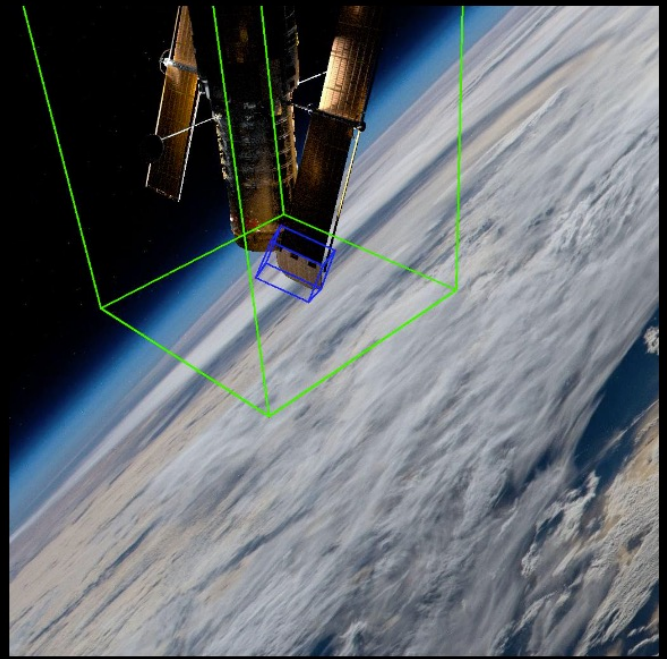


Figure 1: Hubble Space Telescope with earth rendered background, 1024x1024 first query image

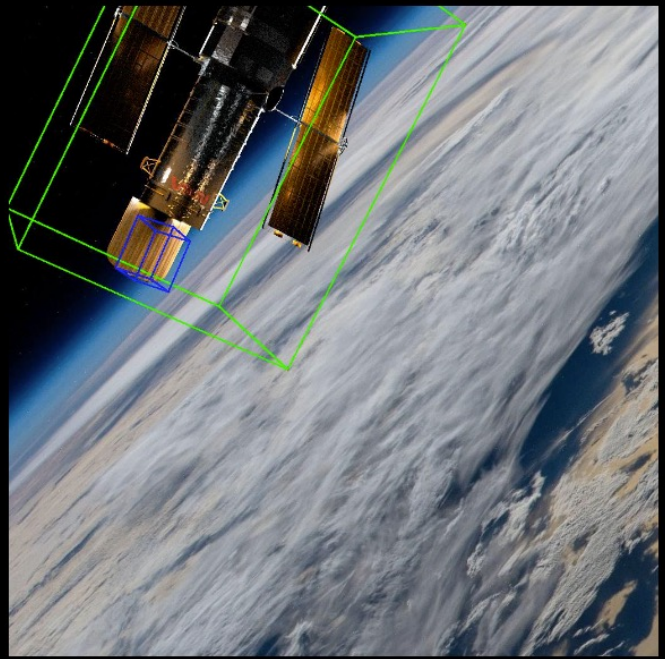


Figure 2: Hubble Space Telescope with earth rendered background, 1024x1024 second query image

Evaluation with all spacecrafts (1)

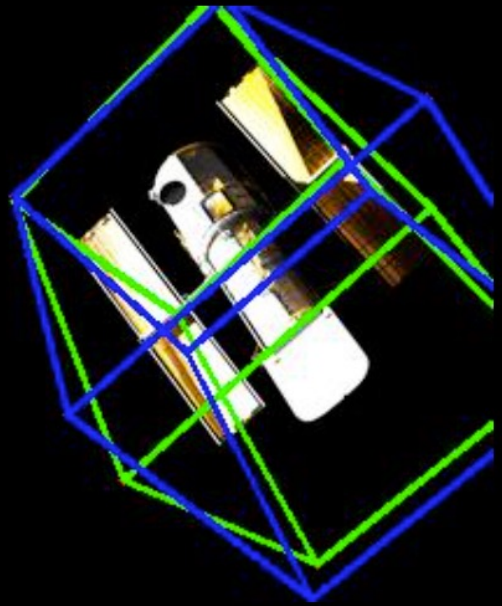


Figure 3: Hubble Space Telescope, no background, 256x256 query image, $e_{\text{ADD}} = 2.925$, $e_{\text{ADD-S}} = 1.183$

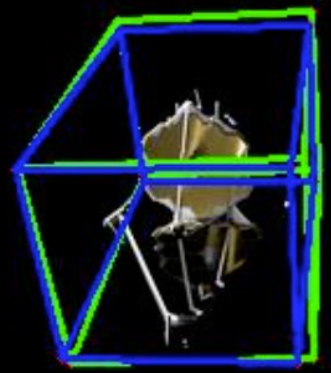


Figure 4: James Webb Space Telescope, no background, 256x256 query image, $e_{\text{ADD}} = 1.415$, $e_{\text{ADD-S}} = 0.808$

Jérémy Chaverot

Evaluation with all spacecrafts (2)

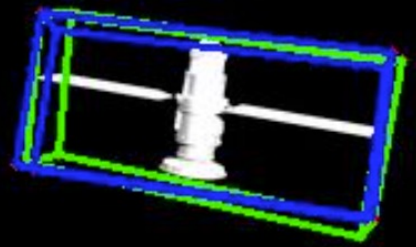


Figure 5: Cosmos Link, no background, 256x256 query image, $e_{\text{ADD}} = 1.718$, $e_{\text{ADD-S}} = 0.383$

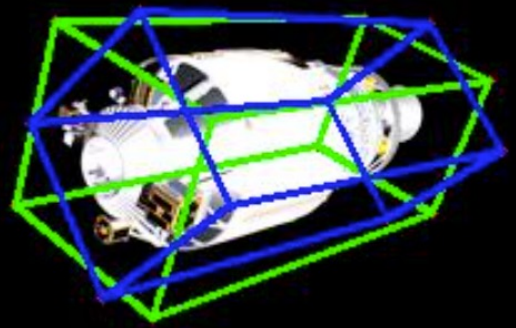


Figure 6: Rocket Body, no background, 256x256 query image, $e_{\text{ADD}} = 1.713$, $e_{\text{ADD-S}} = 0.252$

More results (1)

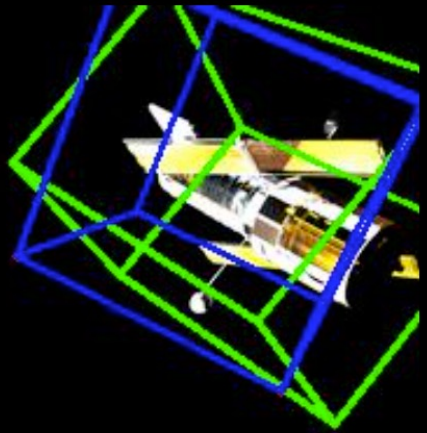


Figure 7: Hubble Space Telescope, no background, 256x256 query image, $e_{\text{ADD}} = 6.514$, $e_{\text{ADD-S}} = 1.571$

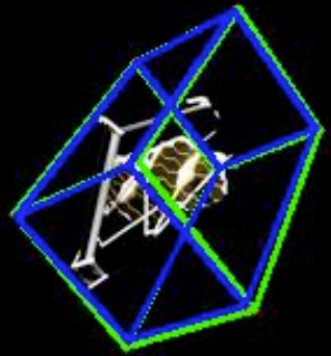


Figure 8: James Webb Space Telescope, no background, 256x256 query image, $e_{\text{ADD}} = 2.224$, $e_{\text{ADD-S}} = 1.261$

Jérémy Chaverot

More results (2)

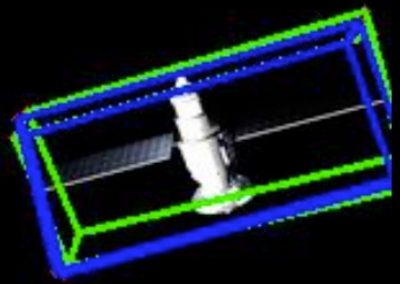


Figure 9: Cosmos Link, no background, 256x256 query image, $e_{\text{ADD}} = 1.925$, $e_{\text{ADD-S}} = 0.377$

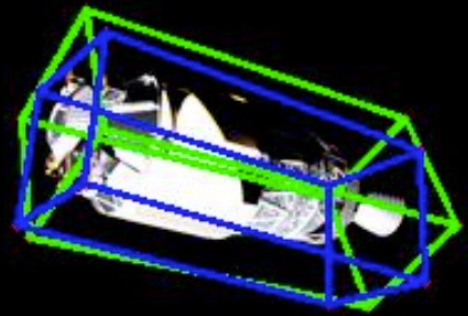


Figure 10: Rocket Body, no background, 256x256 query image, $e_{\text{ADD}} = 1.982$, $e_{\text{ADD-S}} = 0.501$

But mainly poor results (1)

Figure 11: Hubble Space Telescope, no background, intermediary result, $e_{\text{ADD}} = 9.577$, $e_{\text{ADD-S}} = 5.196$

But mainly poor results (2)

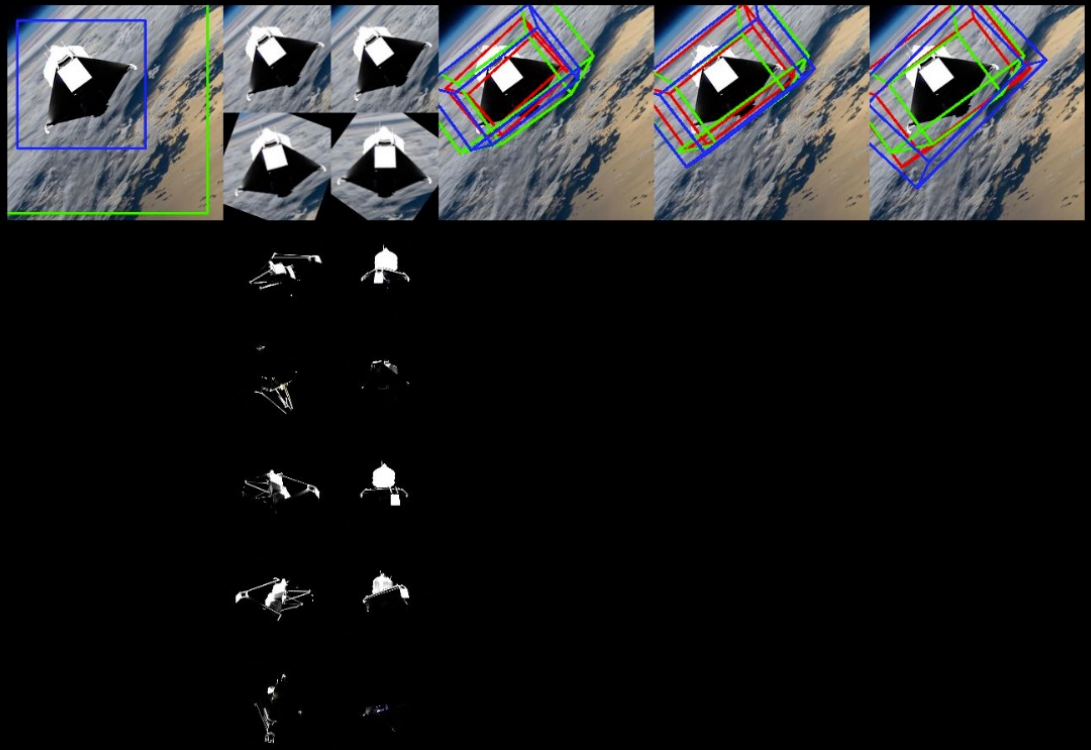


Figure 12: James Webb Space Telescope, with earth rendered background, intermediary result, $e_{\text{ADD}} = 10.934$, $e_{\text{ADD-S}} = 4.317$

But mainly poor results (3)

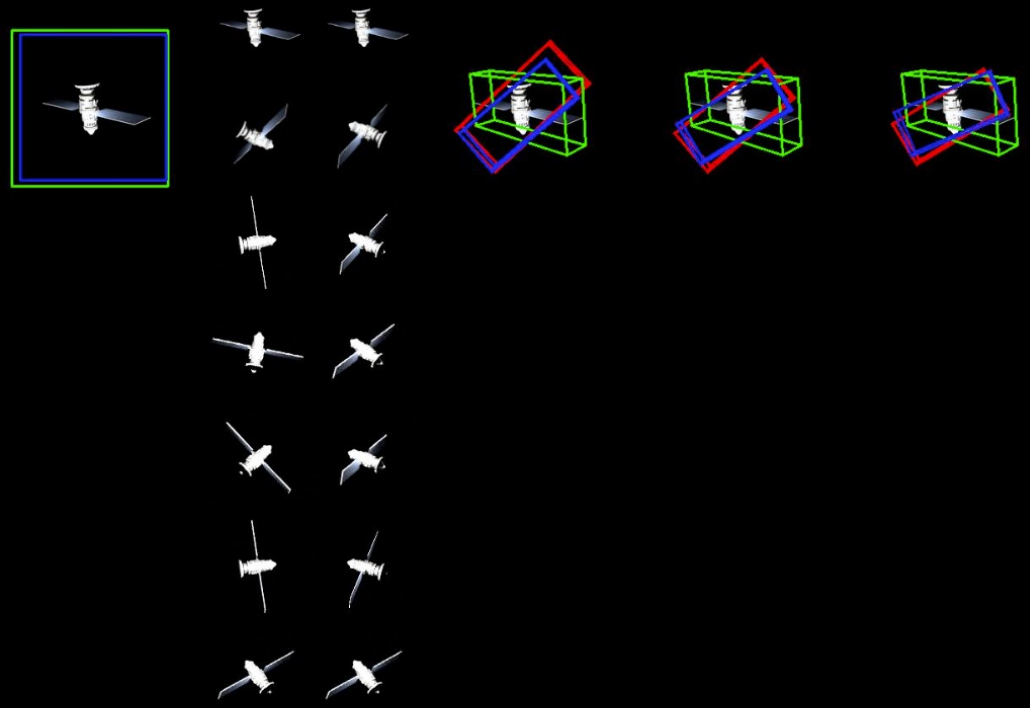


Figure 13: Cosmos Link, no background, intermediary result, $e_{\text{ADD}} = 11.094$, $e_{\text{ADD-S}} = 6.127$

But mainly poor results (4)

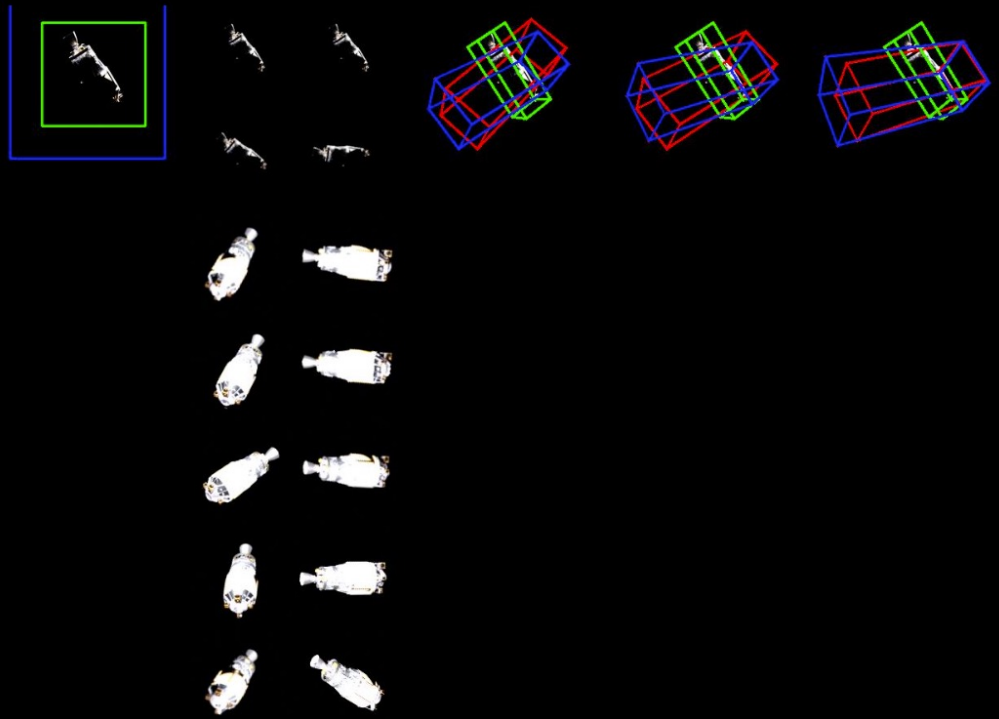


Figure 14: Rocket Body, no background, intermediary result, $e_{\text{ADD}} = 29.335$, $e_{\text{ADD-S}} = 17.743$

Lighting can negatively affect the reference image selection

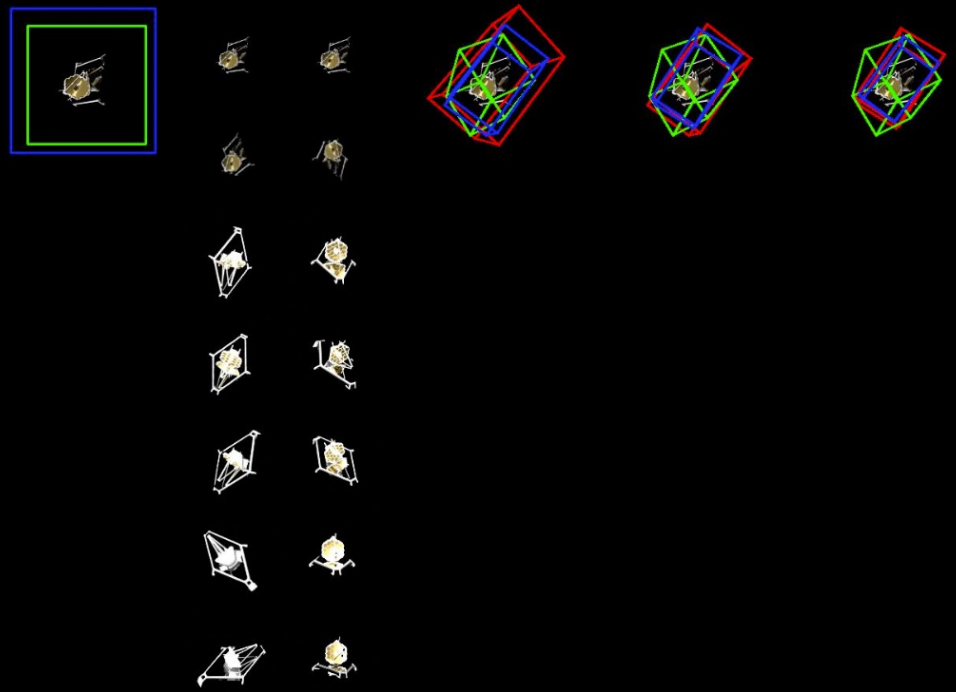


Figure 15: James Webb Space Telescope, with no background, intermediary result, $e_{\text{ADD}} = 21.983$, $e_{\text{ADD-S}} = 12.358$

Coarse pose and refinement are working well

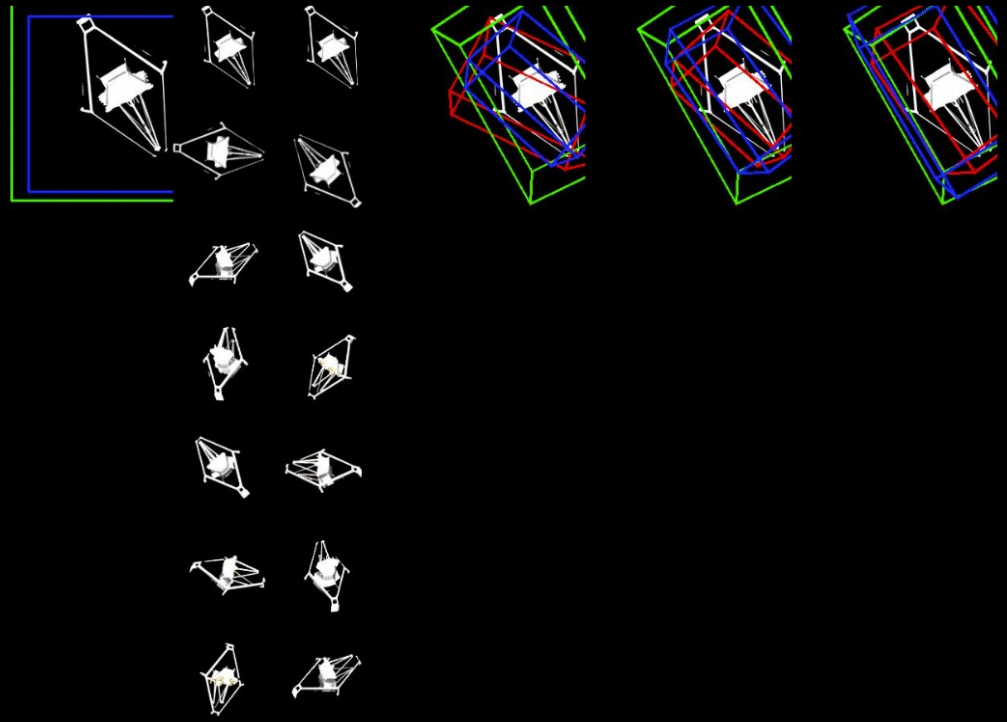


Figure 16: James Webb Space Telescope, with no background, intermediary result, $e_{\text{ADD}} = 1.060$, $e_{\text{ADD-S}} = 0.556$

OUTLINE

Gen6D

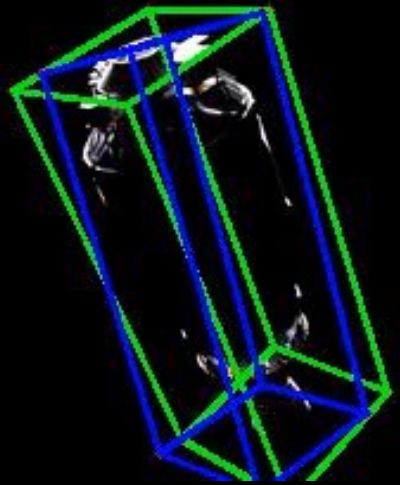
Implementation

Evaluation

Amelioration of the results

Jérémy Chaverot

Tracks for improvements



- Overall the results are poor
- We suggest:
  - Retrain the detector to better handle depth range of spacecrafts (maybe improve → FPN)
  - Retrain the viewpoint selector with various light conditions
  - For the moment, keep the refiner as-is
- Rely more on the 3D model, make use of the segmented images

THANK YOU