Pose Detection For the ESA Pose Estimation Challenge

Introduction

Accurate estimation of distance and orientation (pose) for objects in space is essential for many missions. We have to estimate the pose of the Tango spacecraft from its synthetic and real images captured using computer graphics and a robotic testbed, respectively.

Use cases:

- Remove DEBRIS by Surrey Space Center.
- Restore L by Nasa.
- Phoenix program by DAPRA

Prisma Mission : Launched in 2010, the PRISMA mission demonstrated close proximity operations between two spacecraft in low Earth orbit. Actual space imagery and associated flight dynamics products facilitated the generation of the images used in this challenge.

Dataset : The dataset for this challenge was collected by the Space Rendezvous Laboratory (SLAB), and is part of SLAB's Spacecraft PosE Estimation Dataset (SPEED) benchmark. At SLAB, two key complementary facilities are used to conduct hardware-in-the-loop tests with representative vision-based cameras for space situational awareness and spacecraft proximity operations. Calibrated motion capture cameras report the positions and attitudes of the camera and the Tango spacecraft, which are then used to calculate the "ground truth" pose of Tango with respect to the camera. While these images are used to evaluate the transferability of the submitted algorithms from synthetic to real images, the score calculated on these images is not used for ranking the submissions.

Proposed Approach and Results

Approach is based on a Convolutional Pose machine.

The idea is to extract n key points from the satellite which are predicted from a convolutional neural network. The point in the 3D room of each of these key points is known. After the neural network identifies the key points the image is post processed, the 2D keypoints are extracted and together with the known corresponding 3d coordinates within the satellite model the pose can be estimated with PnP (perspective n-point).

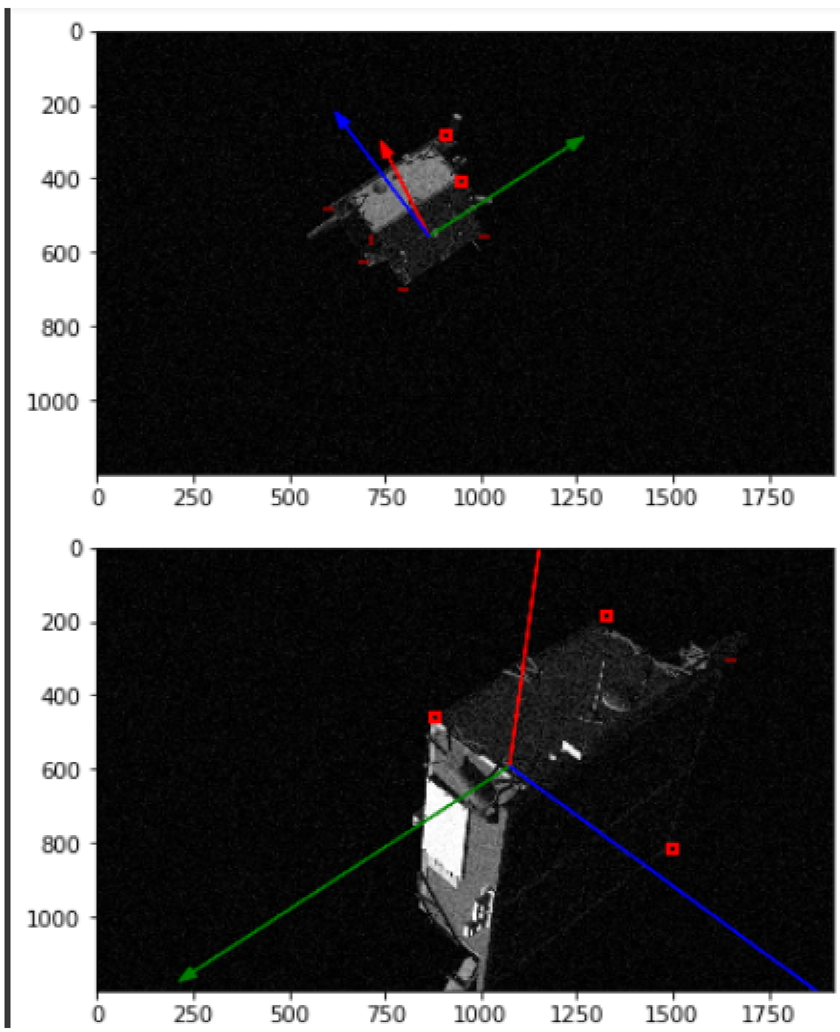
Convolutional Pose Machines(CPMs) inherit the benefits of the pose machine [29] architecture—the implicit learning of long-range dependencies between image and multi-part cues, tight integration between learning and inference, a modular sequential design—and combine them with the advantages afforded by convolutional architectures: the ability to learn feature representations for both image and spatial context directly from data; a differentiable architecture that allows for globally joint training with backpropagation; and the ability to efficiently handle large training datasets.

CPMs consist of a sequence of convolutional networks that repeatedly produce 2D belief maps for the location of each part. At each stage in a CPM, image features and the belief maps produced by the previous stage are used as input. The belief maps provide the subsequent stage an expressive non-parametric encoding of the spatial uncertainty of location for each part, allowing the CPM to learn rich image-dependent spatial models of the relationships between parts.

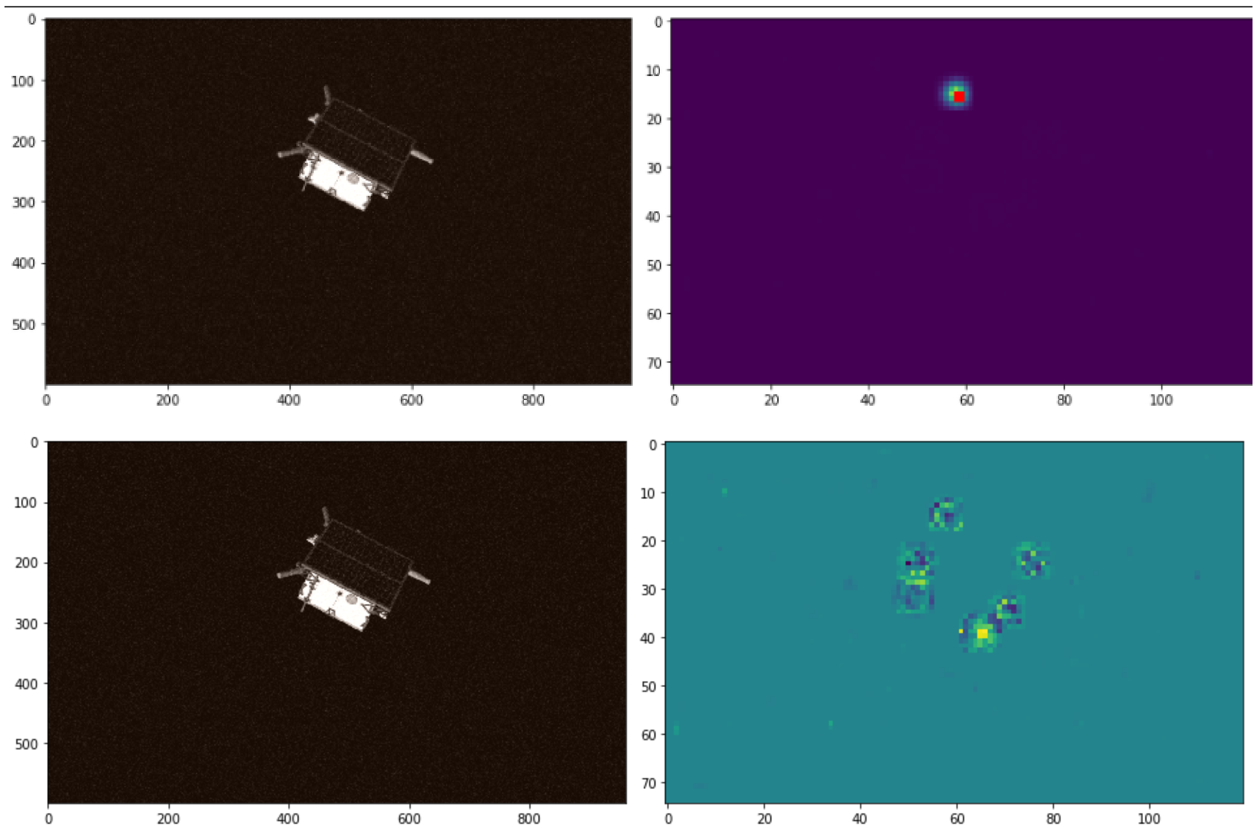
In this approach I do:

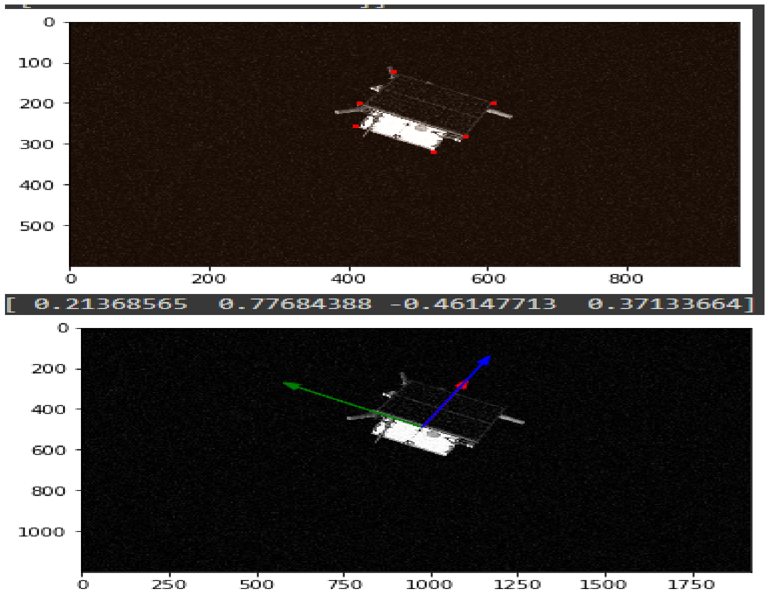
- Create utility Class for accessing camera parameters.
- Computing direction cosine matrix from quaternion, adapted from PyNav.
- Projecting points to the image frame to draw axes.
    1. Determine 8 vertice points.

    2. Determine corresponding 8 surface planes.
    3. Determine 8 vectors between camera and current 3D points(vertitice).
    4. Check if the 8 vectors intersect any of the 8 surface planes. If so, discard points.
- Create Class for dataset inspection : easily accessing single images, and corresponding ground truth pose.
    1. Loading Image as PIL image.
    2. Getting Pose label for image.
    3. Visualizing Image, with ground truth pose with axes projected to training image.
- Visualize images from the Dataset.



- Convolutional Architecture according to https://arxiv.org/abs/1602.00134
- Load trained model for inference
- Data generator definition and declaration.
- Training
- Perform orientation determination and save results.





Future Task to improve Performance

Convolutional pose machines provide an end-to-end architecture for tackling structured prediction problems in computer vision without the need for graphical-model style inference.I showed that a sequential architecture composed of convolutional networks is capable of implicitly learning a spatial model for pose by communicating increasingly refined uncertainty-preserving beliefs between stages. Problems with spatial dependencies between variables arise in multiple domains of computer vision such as semantic image labeling, single image depth prediction and object detection and future work will involve extending our architecture to these problems.