2nd RSEEME Engineering Projects Showcase
ENGN4200 Individual Research Project
May 2019

Application of Stereo Vision for closed-loop reaching control of a robotic manipulator

Student: Nimish Magre
Email: u6434979@anu.edu.au

Primary Supervisor: Dr. Robert Mahony
Secondary Supervisor : Zheyu Zhuang

Scope

Following are the project goals and requirements:

- Experimental Study of monocular vision based network
- New simulator setup for 2nd person stereo camera feed with multiple targets
- Develop CNN architecture to learn the manipulator servo-control parameters
- Set-up rig in lab environment (real-world) and test multiple target grasping success

Previous Work

The project builds on existing research on applying monocular vision for a multiple target reaching task[1]. The research suggests learning a control Lyapunov function (cLf) and its partial derivatives to estimate servo-control parameters for the manipulator using a CNN.

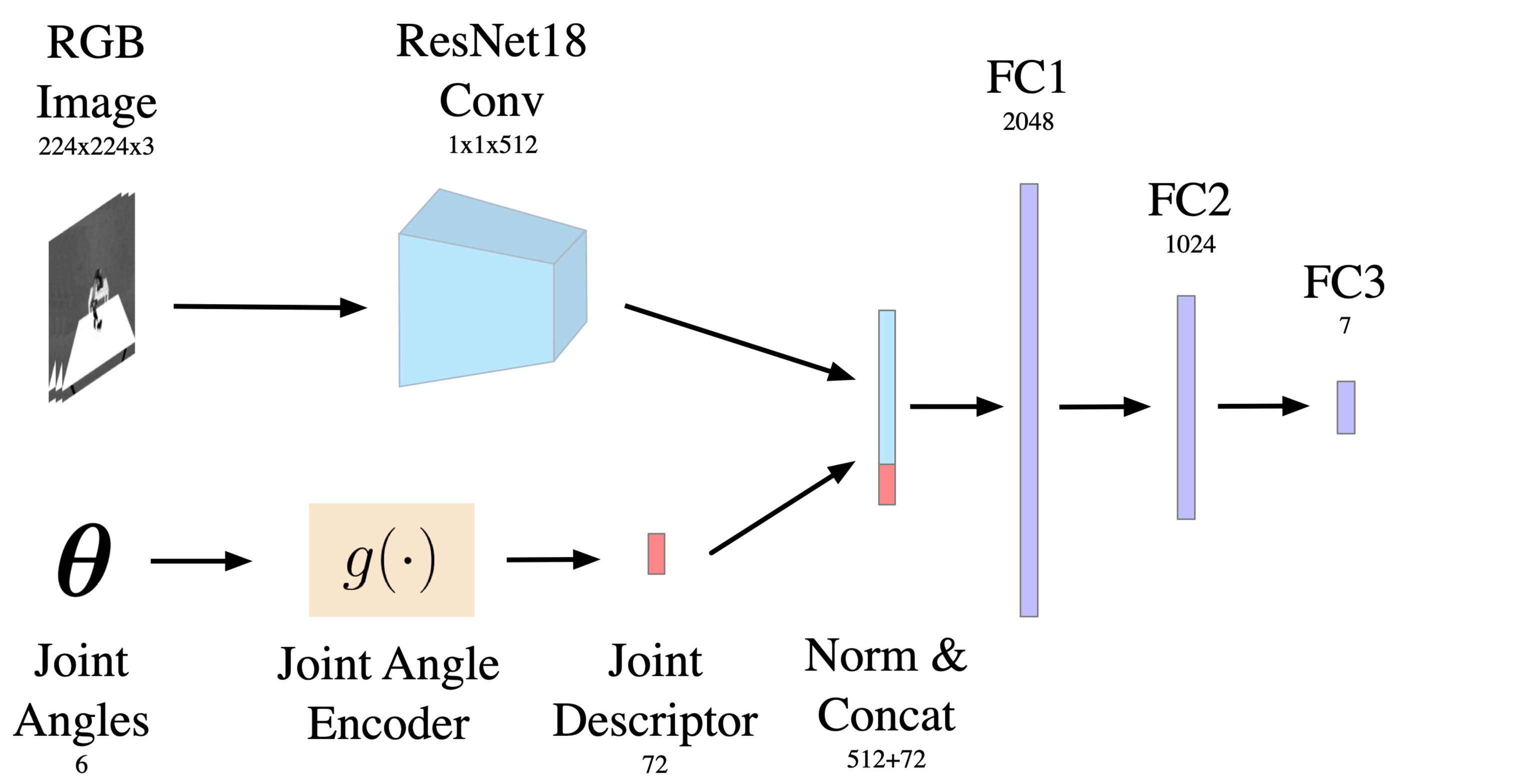

Fig 1 : CNN architecture

It also suggests calculating the cLf with perturbed joint angles separately and comparing it with the original cLf to achieve explicit coupling between the network outputs {learned compound cLf and its partial derivatives}.

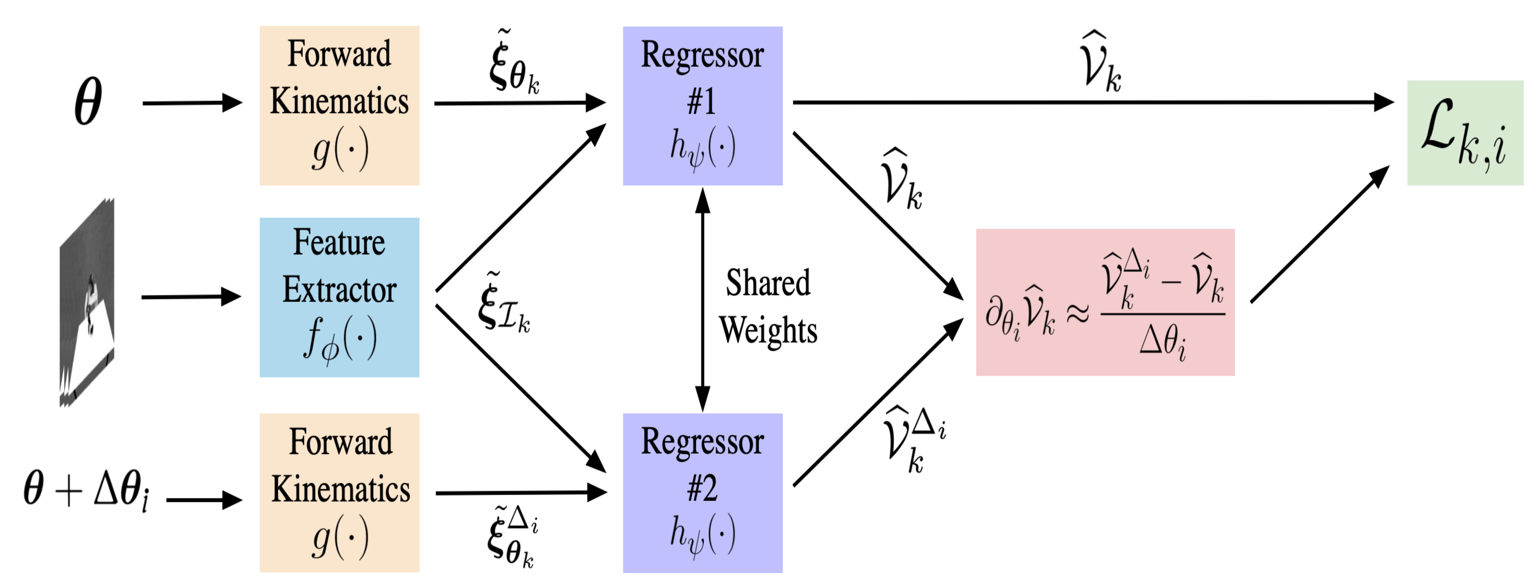

Fig 2: Siamese regressor

Methodology

- Build simulator setup in Gazebo7 and obtain stereo image dataset for training, evaluating and testing the network
- Modify monocular vision based CNN architecture to accommodate stereo images
- Train CNN with generated image dataset from simulator (in-progress)
- Test the trained network with simulator and real-world setup for grasping coloured cube targets

Results

1. **Testing control in real-world setting and in simulation with monocular vision network:**
   Successful grasping rate (real-world) = *78.3 ± 5.4%* [1]
   Successful grasping rate (simulation) = *95 ± 2.28%* [1]

2. **Simulator setup and dataset collection:**

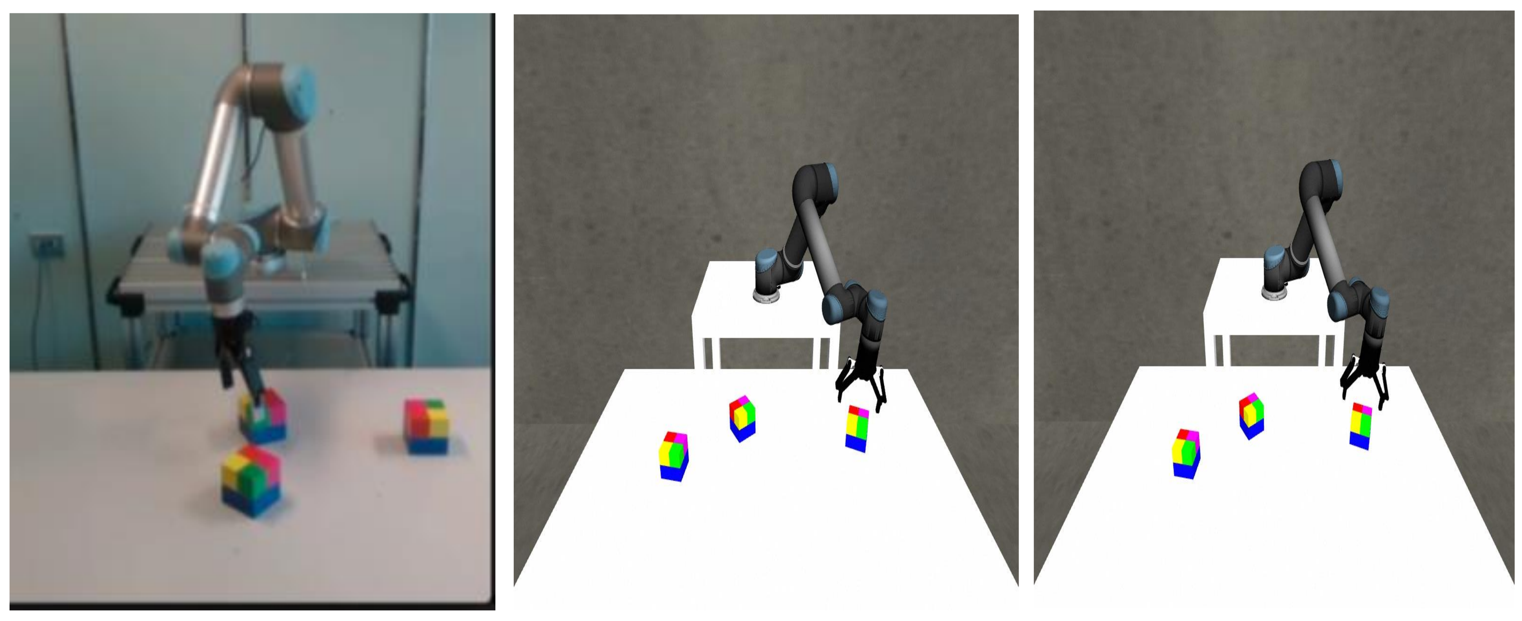

Fig 3: Real-world setup vs left-right stereo images in Gazebo7 simulator

3. **Modified CNN architecture for stereo image pairs:**
   2 parallel image feature extractors with shared weights were used to accommodate the stereo-image dataset [2]

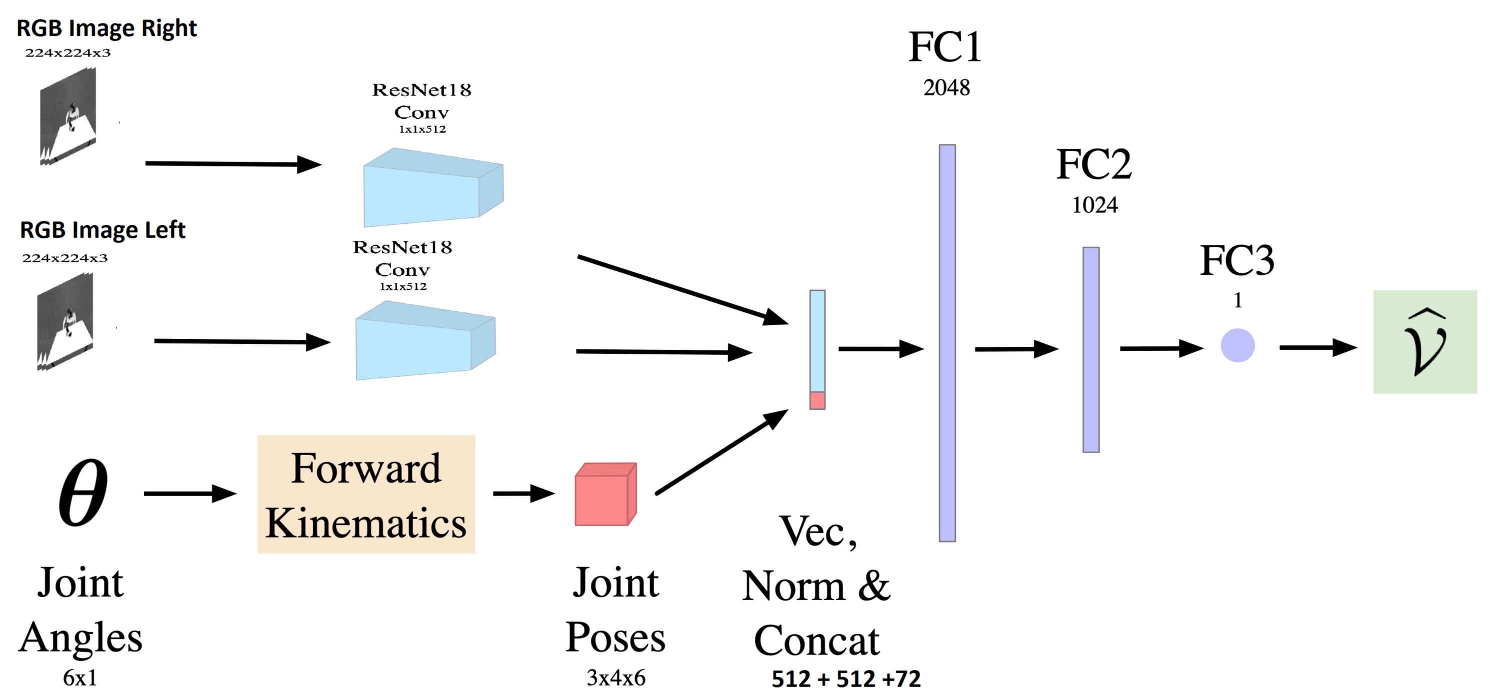

Fig 4 : Modified CNN architecture

Conclusion

When testing the new CNN with 1000 training image pairs, the minimum average training loss plateaued around 32 which is unusual since the network is expected to overfit the small dataset.

I am currently carrying out the following experiments to check for implementation errors

- passing only the right frame images in both the feature extractors
- passing only the right frame images through a single feature extractor and concatenating the features to replicate stereo image features

References

[1] Zhuang Zheyu, Jurgen Leitner, Robert Mahony, "Active Perceptual Cognition: Learning Control Lyapunov Function for Real-time Closed Loop Robotic Reaching from Monocular Vision," Canberra, 2018.

[2] Jia-Ren Chang, Yong-Sheng Chen, "Pyramid Stereo Matching Network," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

Research School of Electrical, Energy and Materials Engineering
ANU College of Engineering & Computer Science