VISION-BASED ROBOTIC GRASP DETECTION USING DEEP LEARNING ALGORITHM

Meriam Joseph

Supervisor: V. Kalaichelvi

BITS Pilani, Dubai Campus, Dubai,UAE

ABSTRACT

Deep Learning has significantly advanced computer vision and natural language processing. It is based on artificial neural networks modeled on the human brain and mimics the way humans learn and remember things. In this project, the deep learning algoritm GR-ConvNet (Generative Residual Convolutional Network) has been implemented to perform optimal vision-based robotic grasping of objects. The project aims to collect image data, train the model using the data and implement the model in real time and perform analysis work.

INTRODUCTION

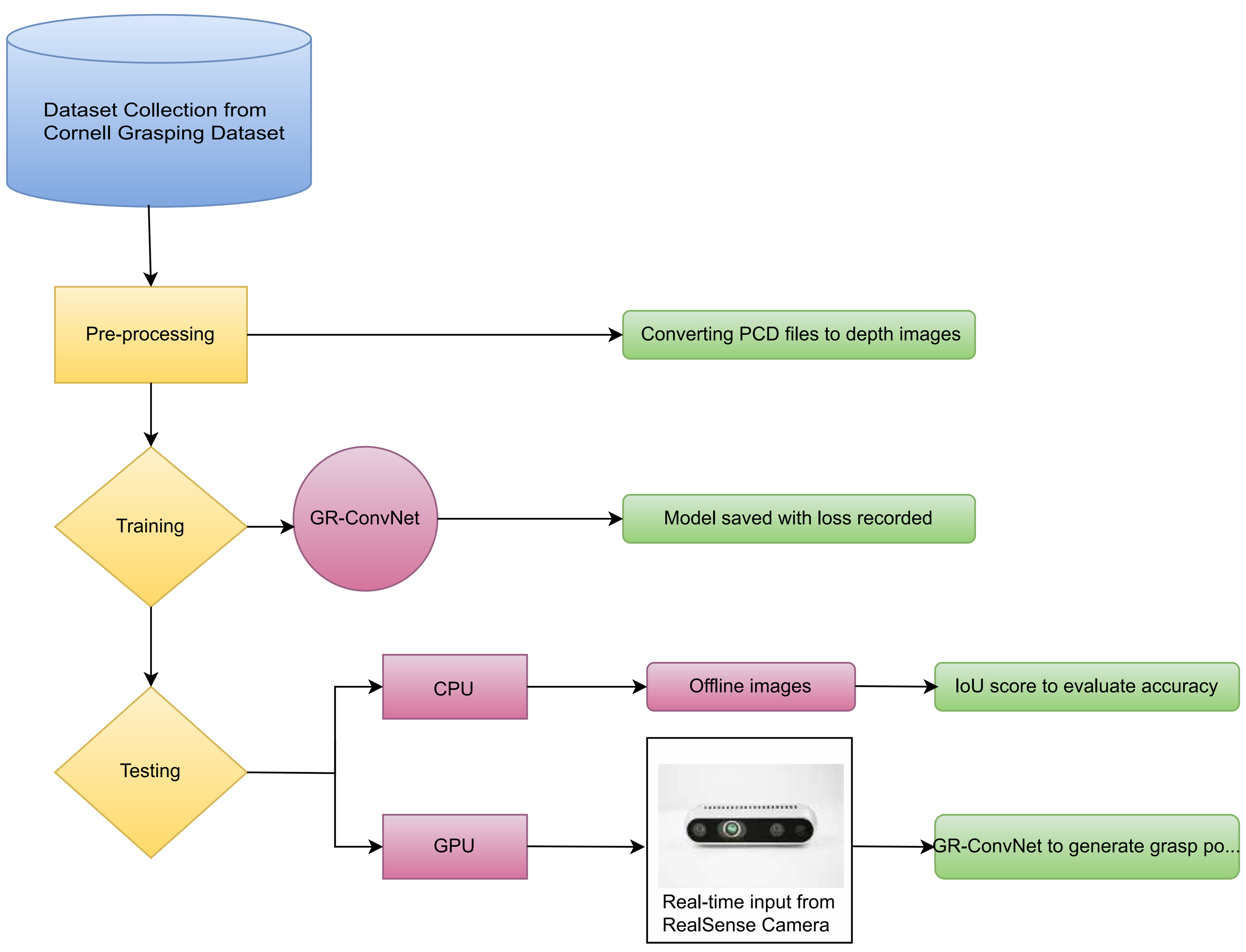
Robotic grasping describes a robot's ability to grasp and manipulate objects in its environment. Humans are extremely capable of manipulation, they can see objects around them and immediately determine the optimal way to pick them up. However, for machines and robots, this is still something that they cannot operate on the same level in. The steadily heavy demand for customization in industries such as mining industries, fashion industries etc. require robots that can perform grasping tasks with the utmost accuracy and precision.

Hence, vision-based robotic grasp detection using deep learning model like GR-ConvNet, aims to achieve optimal robotic grasp predictions for objects. The methodolgy for the same is as follows

1.Collection of data from Cornell grasping dataset.

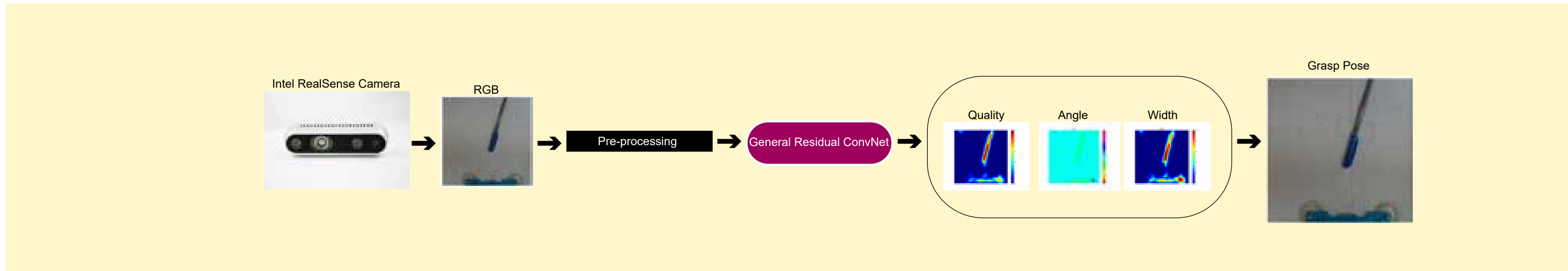
2. Training of data using deep learning GR-Convet model.

3. Real-time testing by using the live image captured from Intel RealSense camera (SR305).

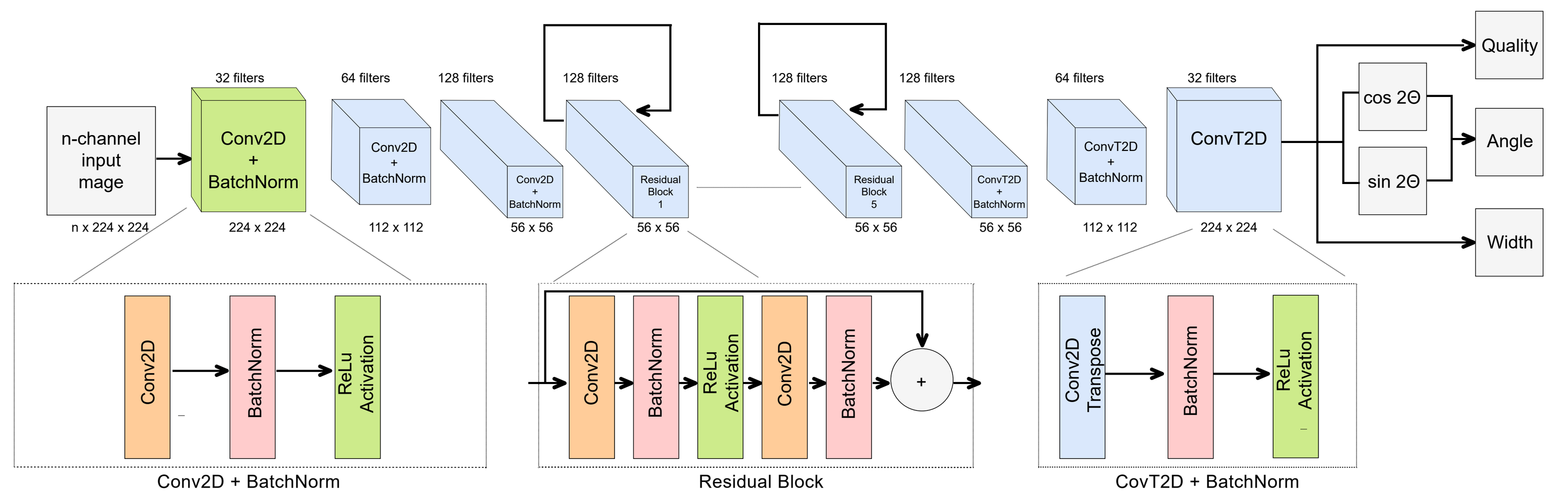


Flow-chart of methodology

ARCHITECTURE DIAGRAM



Inference Model



Architecture Diagram

RESULTS

Training Results

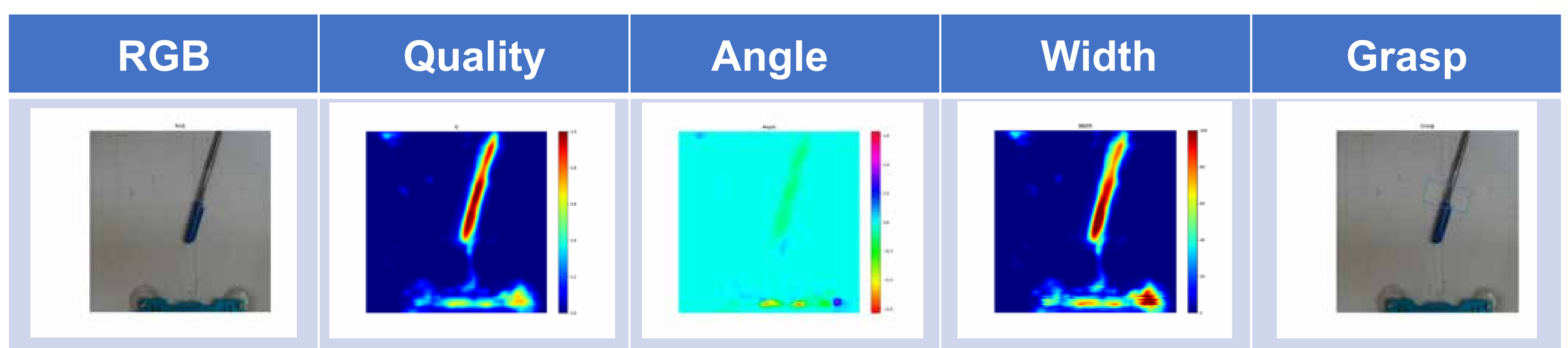
Epoch	Loss	IOU Score
00	0.0841	0.471910 (47.1%)
31	0.0600	0.966292 (96.6%)
49	0.0704	0.865169 (86.5%)

Confidence Threshold	IoU Measure
0.1	0.996292
0.15	0.996292
0.25	0.996292
0.5	0.685395
0.9	0
1	0

Training Results

	CPU	GPU
Average evaluation time per image	1.396 s With data augmentation - 0.887 s	0.086 s
IOU Results	0.966 (96.6%) With data augmentation - 0.943 (94.3%)	0.966 (96.6%)

Evalution model at epoch 31 tested on Cornell dataset with a validation size of 89



REFERENCES

[1] S. Kumra, S. Joshi and F. Sahin, "Antipodal Robotic Grasping using Generative Residual Convolutional Neural Network," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 9626-9633, 2020, doi: 10.1109/IROS45743.2020.9340777.

[2] F. -J. Chu, R. Xu and P. A. Vela, "Real-World Multiobject, Multigrasp Detection," in IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 3355-3362, 2018, doi: 10.1109/LRA.2018.2852777.