Computer Vision Enabled Industrial Robot Manipulator for Sorting Operation

By Ines Haouala

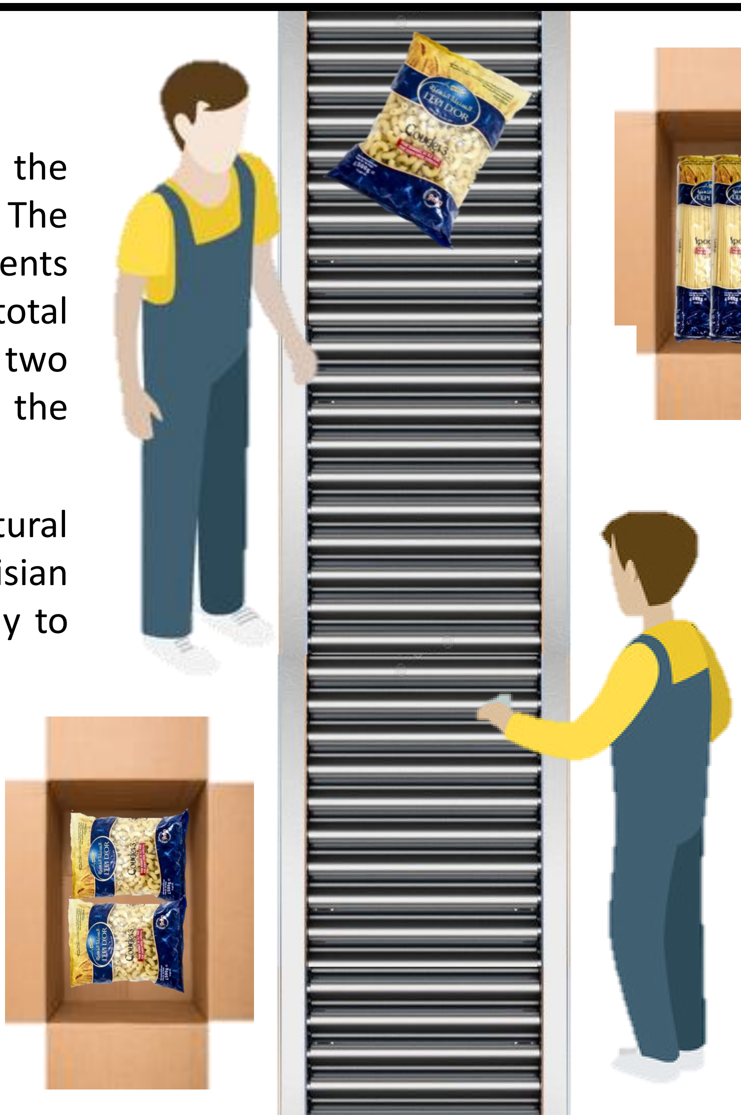
Supervisors: Dr. Charles Kagiri and Dr. Mario Mata

Introduction

In Tunisia, agriculture accounts for 10.4% of the GDP and employs 13% of the workforce. The industry, another prominent sector represents 22.7% of the GDP and employs 32% of the total workforce After the 2011 revolution, these two sectors stagnated and were highly affected by the political instability and past events.

While robots can help skyrocket agricultural production, it is not widely used in the Tunisian agro-industry. This study takes a real case study to improve the efficiency of food handling:

A wheat products supplier is looking into automating his sorting and packaging operation. According to the company, manual sorting and packaging costs them in labor, disturbs and slows down the overall production.



Motivation

The aims are:

- use of industry 4.0 tools to boost the agro-industry in Tunisia
- suggest a new setup of the sorting operation to minimize the work labor and production lead time

The objectives are:

- develop a computer vision-based system to classify pasta types in real-time
- ensure a safe pick and place of the pasta package
- re-create the exact sorting operation on a virtual world using ROS tools
- test the object detector model with real images from the factory sorting operation
- test the ROS generated motion planning of the industrial robot manipulator on a real hardware

Methodology

Working Principle of CNNS

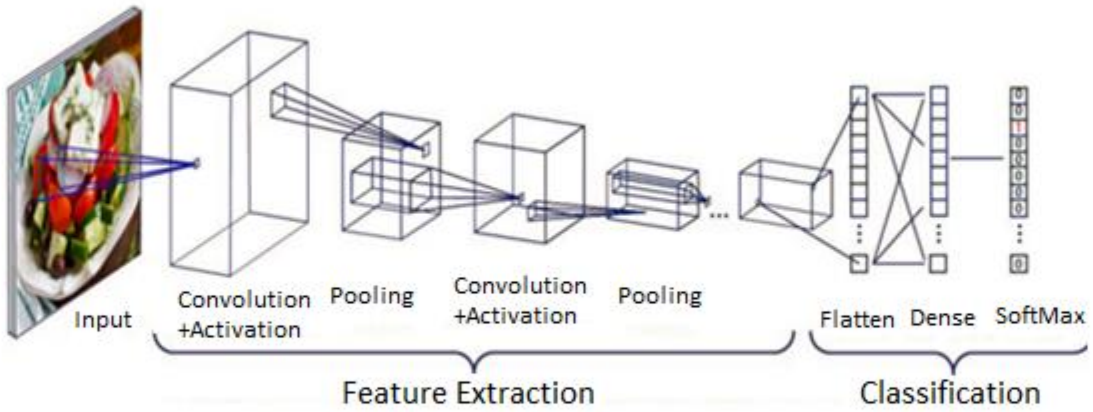


Fig.1: Food classification example using CNN

A fully connected Neural Network or Artificial Neural Network uses a set of neurons and weights to deliver output layers. Whereas, a CNN uses convolutional filters, as seen in in figure 1. Therefore, use of CNNs is lighter, faster and more convenient. Like ANNs, CNNS are difficult to train and take up to 100k images.

Pre-trained CNNS

Model	mAP (%)	GPU(ms)
Darknet53	**78.5**	**26.3**
AlexNet	57	3.1
ResNeXt	70.5	73.8
VGG-16	70.5	9.4

Method	mAP	Time(ms)
YOLO	**33**	**51**
SSD321	28	61
RetinaNet	32.5	73
R-FCN	29.9	85

Data Collection

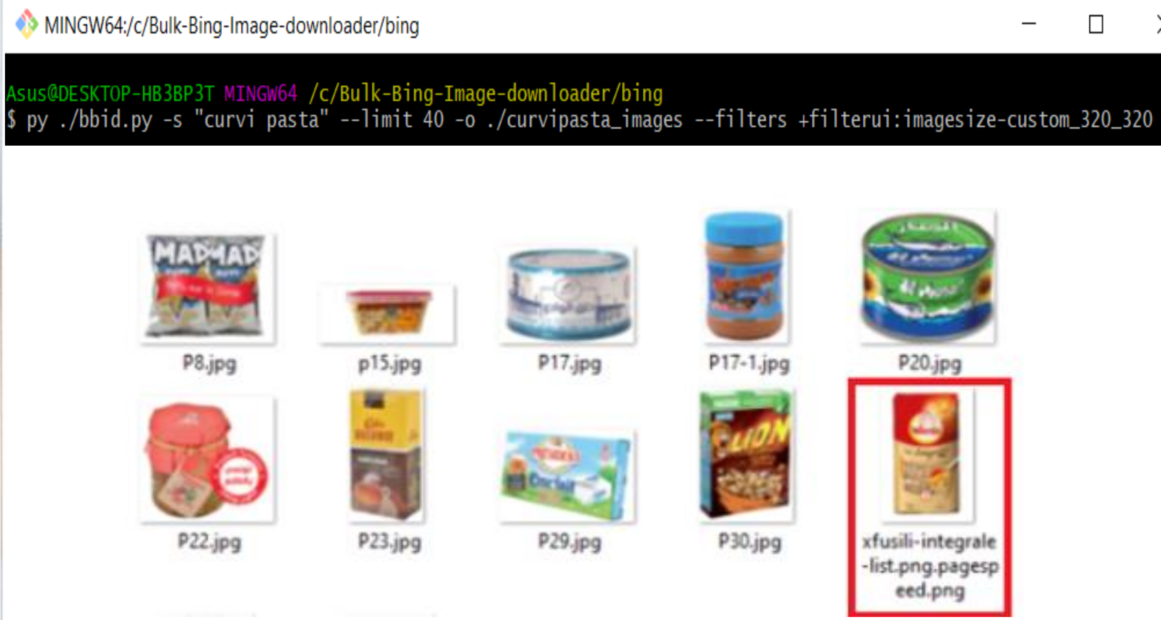


Fig.2: Images scraping with Bulk Bing image downloader

Data Augmentation & Labeling

Fig.3: Augmented image

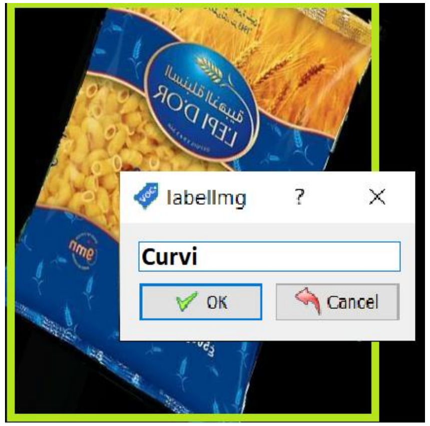


Fig.4: Image labeling

Training on Google Colab GPU

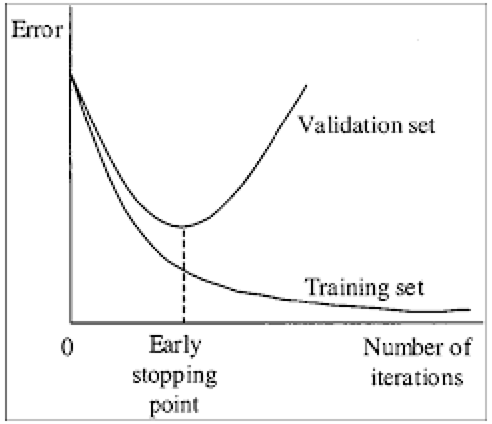
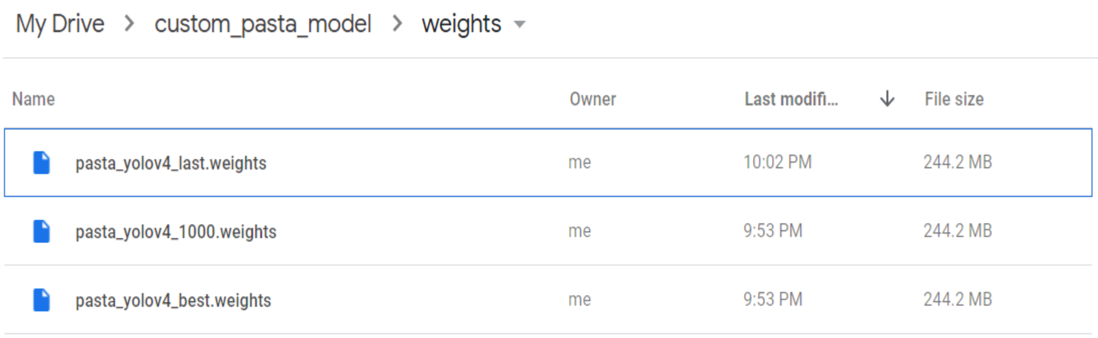


Fig.5: Obtention of best weights



My Drive > custom_pasta_model > weights

Name	Owner	Last modifi...	File size
pasta_yolov4_last.weights	me	10:02 PM	244.2 MB
pasta_yolov4_1000.weights	me	9:53 PM	244.2 MB
pasta_yolov4_best.weights	me	9:53 PM	244.2 MB

Fig.6: Generated weights

Performance Evaluation

$$SPC = \frac{TN}{TN + FP}$$

$$SEN = \frac{TP}{TP + FN}$$

$$PPV = \frac{TP}{TP + FP}$$

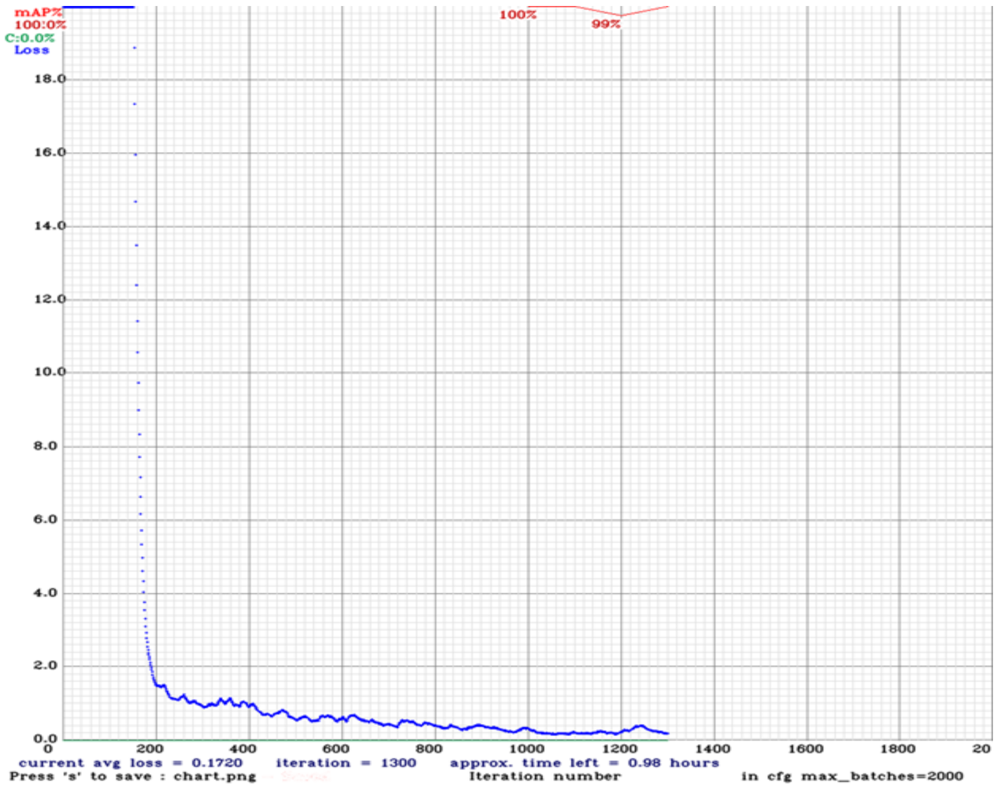


Fig.7: Custom model Loss/mAP chart

Testing image ref	Accuracy in %
1	82.80
2	90.55
3	81.03
4	92.19
5	94.91
6	92.05

Testing image ref	Accuracy in %
7	80.37
8	90.08
9	95.17
10	96.75
11	74.08
12	97.42

Testing image ref	Accuracy in %
13	94.30
14	98.44
15	95.98
16	96.93
17	97.36
18	96.27

Pose Estimation and 3D Reconstruction

Fig.8: Stationary mounting of RGB-D camera

$$XYZ_{camera} \xrightarrow{HAND-EYE} XYZ_{EE} \xrightarrow{ROBOT-POSE} XYZ_{robot}$$

$$A_{ij}X = XB_{ij}$$

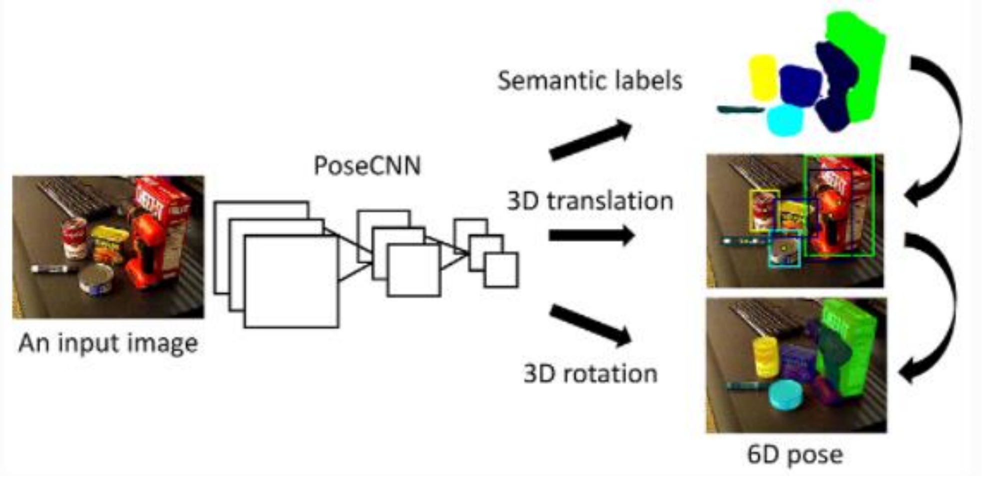


Fig.9: Trained CNN for 6D pose estimation

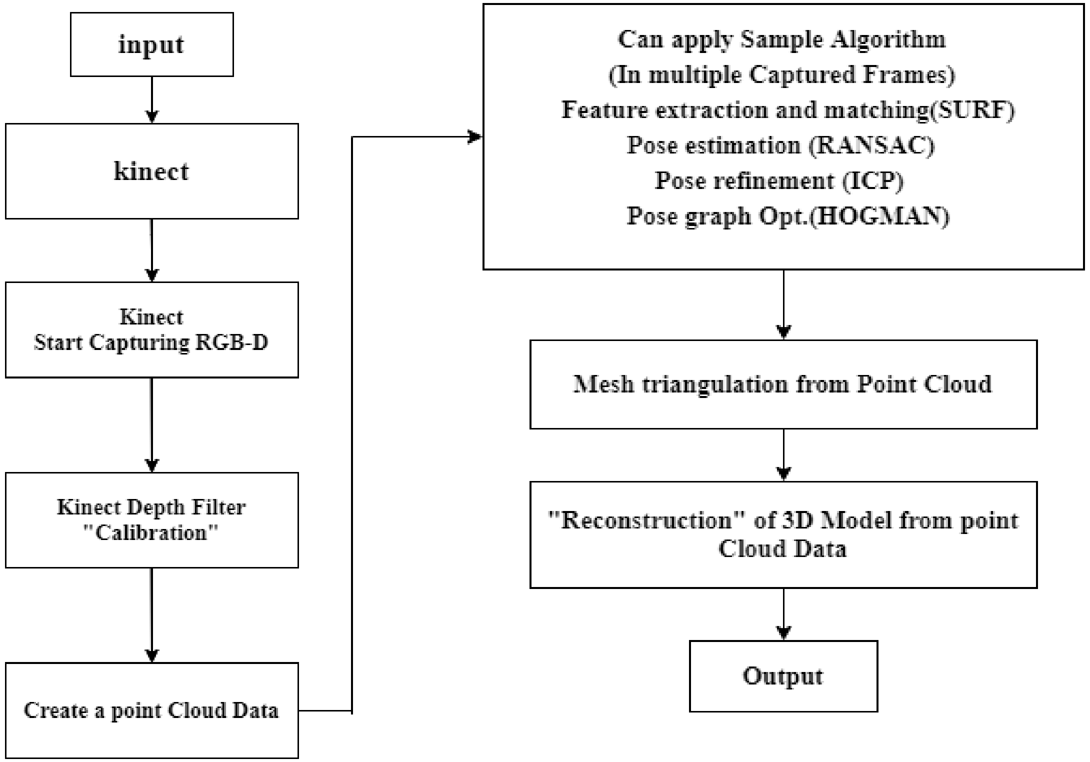


Fig.10: Camera 3D reconstruction flowchart

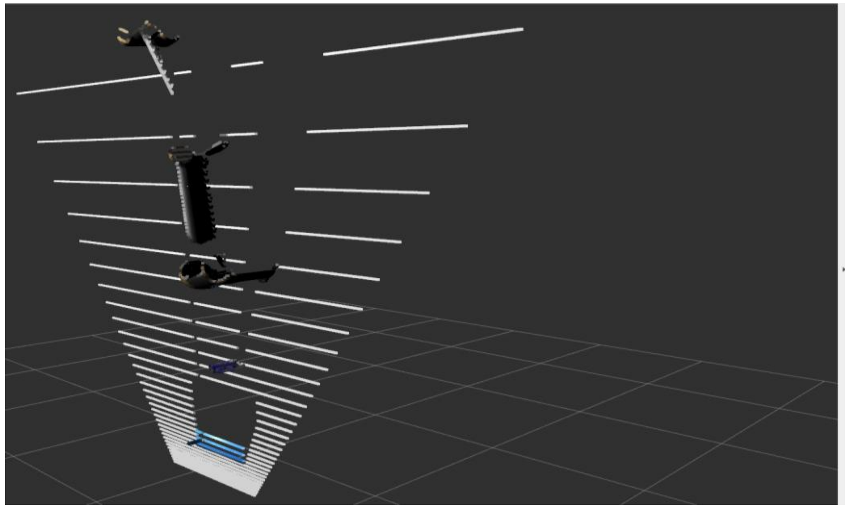
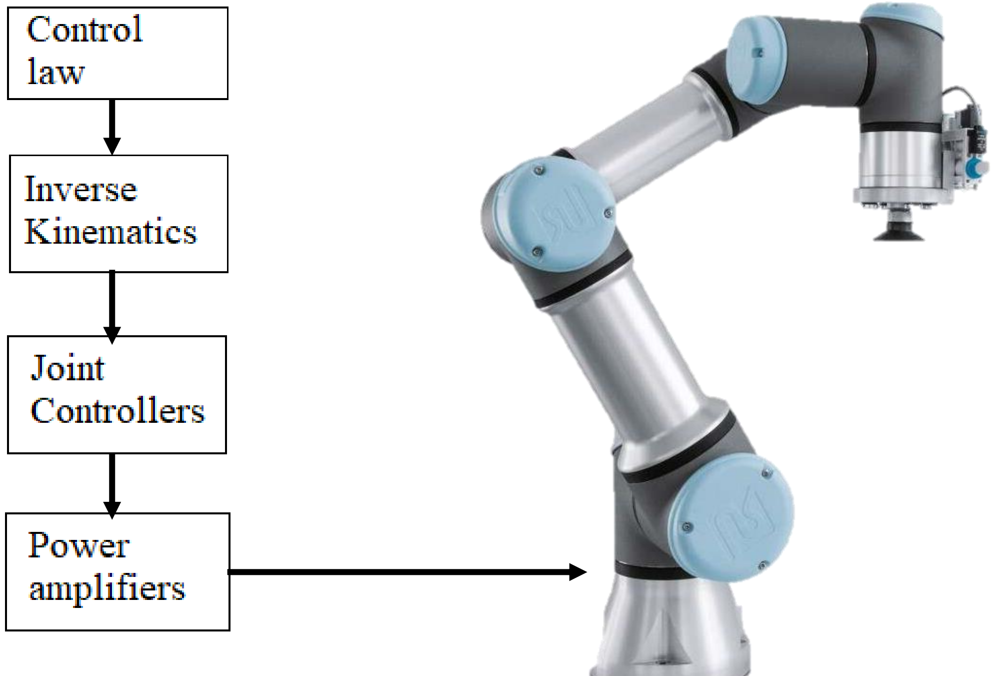


Fig.11:Simulation of Kinect output on Rviz

Industrial Robot Manipulator

The final selection of industrial robot manipulator was based on the following criteria: application, payload versus cost, degree of freedom, maximum operating range and end-of-arm tooling. SCARA was found to be the cheapest and most convenient option, however, for this project Universal Robot UR3 was used for simulation purposes.

$$A_i = \begin{bmatrix} \cos\theta i & -\sin\theta i \cos\alpha_i & \sin\theta_i \sin\alpha_i & l_i \cos\theta i \\ \sin\theta i & \cos\theta i \cos\alpha_i & -\cos\theta i \sin\alpha_i & l_i \sin\alpha_i \\ 0 & \sin\alpha_i & \cos\alpha_i & di \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



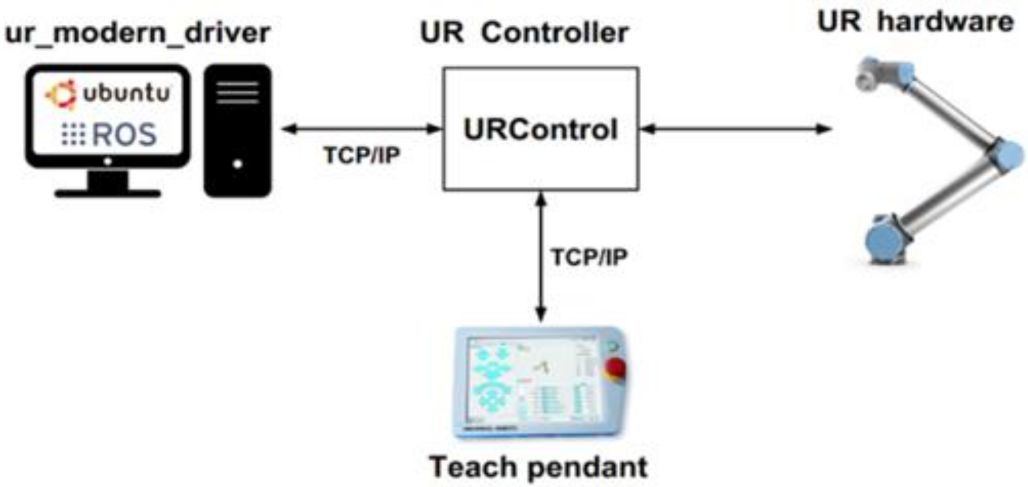


Fig.12: UR3 control system architecture

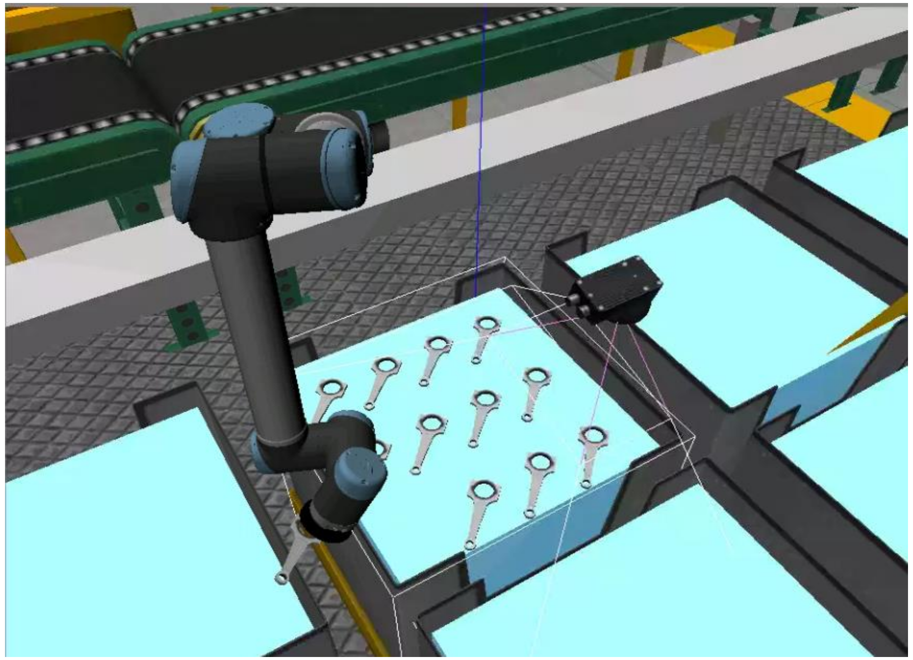


Fig.13: Motion planning using ROS tools

Conclusion

Key Findings

- UR3 and Kinect camera are the most convenient solutions, performance, and cost-wise, for the sorting operation
- Training a pre-trained CNN Darknet + yolov4 with 400 images dataset is relatively fast and presents high accuracy, with no less than 70% of accuracy on all validation images
- The custom CNN also returned FP when tested with more complex images, therefore the model is underfitting
- Pose estimation is essential to ensure safe pick & place of the package
- ROS stood out as an open-source framework that provides the tools and libraries to create complex robot applications
- Results are promising to acknowledge the contribution of industry 4.0 technologies to the agro-industry field

Future Work

- Train with a more extensive and more diverse dataset to improve the model flexibility and performance of the custom object detector/classifier
- Test the object detector model with real images from the factory sorting operation
- Export the custom object detector/classifier to a ROS node
- Test the ROS generated motion planning of the industrial robot manipulator on a real hardware
- Re-create the exact sorting operation on a virtual world using ROS tools
- Use of visual servoing technique for the control of a robot manipulator