# An Undergraduate R&D Internship/Project on "Machine vision based categorical quantization"

By


**Md Zafore Sadek**

Student ID: **1720525**


**Spring, 2022**

At


Supervisor:

**Ajmiri Sabrina Khan**

**Lecturer**

Department of Computer Science & Engineering

Independent University, Bangladesh


**May 11, 2022**

**Dissertation submitted in partial fulfillment for the degree of Bachelor of Science in Computer Science**

**Department of Computer Science & Engineering**

**Independent University, Bangladesh**

# Attestation

I, Md Zafore Sadek, hereby certify that none of the work that has been done in this report is plagiarized. Any resources used are mentioned in the bibliography section of the report. This report is submitted in fulfillment of the internship requirement for the Bachelor of Science in Computer Science and Engineering degree at Independent University, Bangladesh. It has been constructed to serve as documentation for the professional internship experience that I was engaged in at Animo.AI during the Spring 2022 session. The only parties that have contributed in guidance of constructing this report are my internal and external supervisors at IUB and Animo.AI.

Signature

Date

Md Zafore Sadek

Name

# Acknowledgement

I am grateful to Allah for keeping me safe in this pandemic and giving me the patience to work with my knowledge.

I would like to express my gratitude to my internal supervisor Ajmiri Sabrina Khan, Lecturer, Independent University, Bangladesh. She has provided me with the guidance and suggestions necessary to conclude all research, industrial and documentation based tasks associated with my internship at Animo.AI.

I would also like to thank my external supervisor Mostahid Ahmed, Chief Executive Officer, Animo.AI, for granting me the opportunity to contribute in the state of the art projects at the company and learn from this experience.

Lastly, I appreciate the authorities and regulatory bodies at Independent University, Bangladesh that are conducting the internship program and validating our internship experience through constructive criticism.

Md Zafore Sadek
1720525
May 2022
Dhaka, Bangladesh

# Letter of Transmittal

May 11, 2022
Ajmiri Sabrina Khan
Internal Internship Supervisor & Lecturer
Department of Computer Science and Engineering
School of Engineering, Technology and Sciences
Independent University, Bangladesh

Subject: Documentation of my work progression on "Machine vision based categorical quantization", a customize-able Industry 4.0 service developed and deployed by our team at Animo.AI.

Greetings,

It is a pleasure to share this documentation, it includes a record my work and contribution in the "Machine vision for categorical quantization" project at Animo.AI. The agenda of this report is to discuss the activities that i was involved with in the internship period and serve as fulfillment of the internship requirement for the Bachelor of Science in Computer Science and Engineering degree at Independent University, Bangladesh.

I would like to take this opportunity to express my gratitude for your time, expertise, guidance and support. I have made a notable effort in the preparation of this report and I hope that the purpose of this report is served successfully. Lastly, I am open to any relative queries.

Thanks & Best Regards
Md Zafore Sadek
1720525

# Evaluation Committee

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Signature

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Name

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Supervisor

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Signature

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Name

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Internal Examiner

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Signature

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Name

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

External Examiner

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Signature

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Name

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

Convener

# Abstract

The purpose of this report is to accumulate the details of my experience while working with the "Computer vision for categorical product quantization" project at Animo.AI.

My work assignment included leveraging technologies like artificial intelligence, computer vision and deep learning to develop and deploy an automated system that performs categorical product quantization by utilizing the already existent surveillance cameras in an industrial environment.

The tasks that I attended to in this project pipeline consisted of data set creation and preparation, deep learning model selection and configuration, industrial environment analysis with the goal to identify performance factors and achieve industry grade performance.

Throughout this experience, I have gained knowledge of how data oriented intelligent systems should be developed and deployed in an industrial environment where there is no scope of trade off in performance.

Animo.AI is focused on implementing new ways of analyzing data by leveraging Machine Learning technologies to develop high-end Industry 4.0 solutions for industries and enterprises to generate business value through data.

This report concludes itself through discussion on the problems faced during the implementation of the project, the performance that the current version of the project provides after handling the problems that had arisen during implementation and the future scope of this project.

***Keywords***— Artificial Intelligence, Machine Learning, Deep Learning, Computer Vision Industry 4.0

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview/Background of the Work

Animo.AI is a start up company focused on implementing new ways of analyzing data by leveraging intelligent state of the art technologies to develop high-end Industry 4.0 solutions for industries and enterprises to generate business value through data.
This project was initiated as a proof of concept (POC) project when Animo.AI started its agreement with XYZ Industries Limited. Here, the actual identity of the industry cannot be revealed due to non disclosure policies at Animo. XYZ Industries Limited is the largest manufacturer, distributor and marketer of biscuits in Bangladesh. The requirements for this project were specifically outlined by XYZ Industries Limited. The problem statement that the client addressed goes as follows; The necessary activity of categorically counting product packages before they are moved to the logistics department for country wide distribution is undergone in a manual process. This requires massive human labour. The purpose of this project was to move forward with a computer vision approach and utilize the already existent surveillance assets(CCTV) to automate the activity of categorically counting product packages.

## 1.2 Objective

The objective of this project was to prove the concept that the activity of categorically count product packages with an automated computer vision approach is plausible and can deliver industry grade performance.

## 1.3 Scope

The asserted scope was to fulfill the project objective within my internship period. This would enable Animo.AI to move forward with full scale deployment as soon as the POC project would pass the accreditation tests done by the authorities at XYZ Industries Limited.


Figure 1.1: Example view from a camera mounted above a conveyor belt.

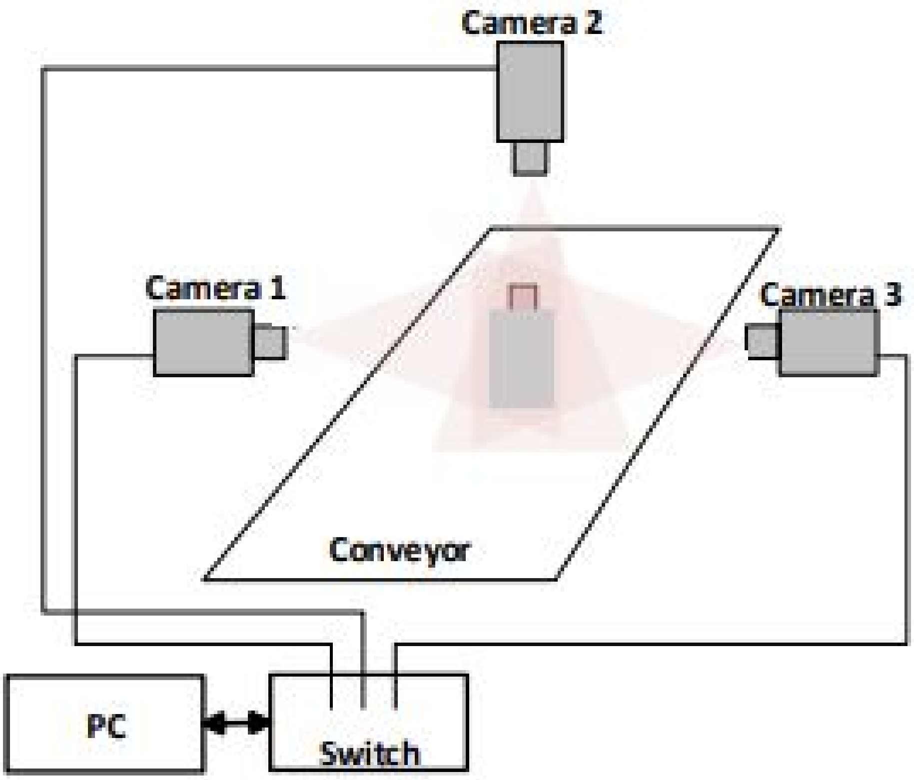


Figure 1.2: General structure of machine vision systems.[1]

# Chapter 2

# Literature Review

## 2.1 Relationship with Undergraduate Studies

Independent University, Bangladesh offers advanced courses in the domain of Intelligent Visual Computing that aided in the development and deployment of the POC project. The courses are as follows:

- CSE 317, Numerical Methods : This course introduced me to the modern numerical approximation techniques necessary for in-depth data analysis.
- CSE 417, Data Mining and Warehouse : This course is a combination of data analysis techniques such as cleaning, transforming, warehousing, cubing, classification, clustering and mining frequent patterns
- CSE 420, Image Processing : In this course topics include point operations, histograms, spatial operations, image rectification, interpolation and other transformations, contrast enhancement, magnification, Fourier image transforms, edge and contour detection, boundary extraction and representation. Focus is mainly placed on the general principles of image processing. Other topics discussed include morphological image processing, wavelets, compression and convolution operations for the tasks of image classification, localization and detection.
- CSE 421, Machine Learning : Supervised learning: Information theoretic decision tree learner, best current hypothesis search, candidate elimination (version space) algorithm, learning in the first order Horn clause representation, inductive logic programming, applications; Unsupervised learning: hierarchical clustering, category utility, incremental and non-incremental algorithms for hierarchical clustering, applications; Connectionist learning: introduction to neural networks, Feed forward and recurrent networks, perception, multilayer feed forward networks, backpropagation algorithm for training a feed forward network, applications; Genetic algorithms: genetic operators, fitness function, genetic algorithm in supervised learning framework, applications.

- CSE 424, Neural Networks : This course included lectures and exercisez on image classification pipeline, loss functions and optimization, convolutional networks, implementation of 2D convolution and maxpool from scratch, class activation maps pointing towards the visual attention of the model, object detection and localization, image segmentation, atrous convolution for semantic segmentation, encoders and decoders, sequential modeling with RNN and LSTM, image captioning using attention, word vectorization, graph convolutional networks, node embeddings and classification.

- CSE 425, Artificial Intelligence : This unit covers the foundational concepts and programming techniques of AI: search and problem solving methods, knowledge representation, reasoning, intelligent agents and natural language processing. Additional aspects of AI discussed include logic, uncertainty, puzzle solvers, simulative and cognitive process, expert systems and data processing.

## 2.2 Related works

- An Image Processing based Object Counting Approach for Machine Vision Application[2]: Machine vision applications are low cost and high precision measurement systems which are frequently used in production lines. With these systems that provide contactless control and measurement, production facilities are able to reach high production numbers without errors. Machine vision operations such as product counting, error control, dimension measurement can be performed through a camera. In this paper, a machine vision application is proposed, which can perform object-independent product counting. The proposed approach is based on Otsu thresholding and Hough transformation and performs automatic counting independently of product type and color. Basically one camera is used in the system. Through this camera, an image of the products passing through a conveyor is taken and various image processing algorithms are applied to these images. In this approach using images obtained from a real experimental setup, a real-time machine vision application was installed. As a result of the experimental studies performed, it has been determined that the proposed approach gives fast, accurate and reliable results.

- Vision based real-time fish counting, inspection and classification using deep learning[3]: Current fish counters rely on feeding fish through dedicated equipment as a part of the fish transportation. This thesis proposes a vision-based alternative, using cameras mounted above conveyor belts to count, inspect, and classify fish. The proposed solution is based on a multiple object tracking algorithm, using deep learning to detect and track fish from frame to frame in a video. Tracking of fish through a video ensures that each fish is only counted once, and it also enables fish inspection and classification. Thus, in addition to fish counting, this thesis also investigates damage detection approaches and methods for classifying fish as dead or alive.

# Chapter 3

# Methodology

## 3.1 Agile Approach

This project was conducted in an agile process. Agile processes are incremental in nature. It is conducted in 1/2 week sprints. Here, each sprint begins with defining a backlog of activities, the completion of which should provide an output close to the goal of the project. After each sprint the outcomes are reviewed and a new backlog of activities is defined to reach a better outcome then the one in the first sprint. This is a repetitive cycle which ensures periodic improvement of the project outcomes.

## 3.2 Data Preparation

Data integrity is a very important factor at XYZ Industries Limited, thus, the main server at their industry is a local server. Periodic feed from the camera mounted on the conveyor belt would be stored in a computer at their server room. This feed would then be sent to us. We would split the feed in two parts; the training part and the validation part. The feed in the validation part would be divided further into multiple spill videos, all of which would be used for validation after training. The count of each category in the spill videos would be manually confirmed before validation process. The training feed was converted to frames in a manner where each training frame was selected at intervals of 8 frames. This would result in the accumulation of 3/4 training frames from each second of the feed. Each of the objects that required categorical quantization in these training frames was then annotated with the image annotation tool LabelImg. An object is identified with its bounding box (bbox) and the associated annotation is its class name and its coordinates in the image.

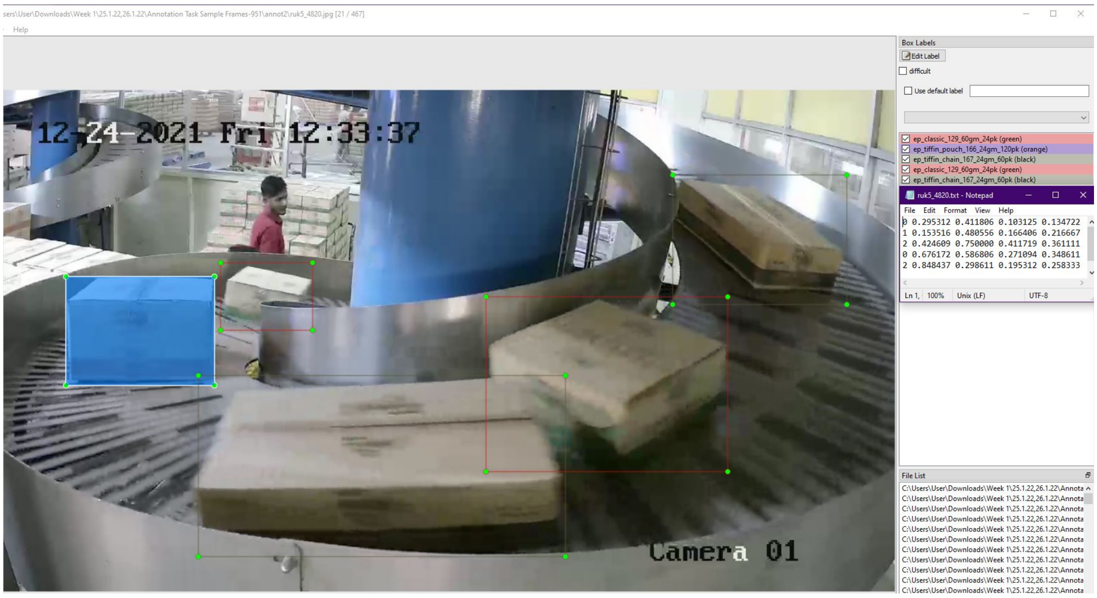


Figure 3.1: Annotating training frames with LabelImg

## 3.3 Multiple Object Tracking

Multiple object tracking (MOT) is a computer vision problem that attempts to identify and track multiple objects in a video sequence, keeping track of their positions and trajectories. Each object is tracked in a track, which contains a unique ID and information of the object from previous frames. The most commonly used strategy, and the one used in this project, is detection-based tracking (or "tracking-by-detection") [6]. This method consists of four main parts or stages for each frame of the video. First is the detection stage, where all the objects in the frame are identified. This is followed by a motion prediction or feature extraction stage, where the goal is to either predict the position of existing tracks, or to extract features such as appearance features of the objects. This is followed by an affinity / cost stage, where all the objects in the new frame are compared to all the existing tracks, and given an affinity or cost score based on a chosen metric, such as distance or appearance similarity. Lastly, a matching algorithm is used in an association stage to match detected objects to existing tracks, and handle the birth/death of tracked objects [4]. Within detection-based tracking there are different models that can be used, such as an appearance model or a motion model. The appearance model uses the visual representation of an object to calculate similarity between objects, while the motion model uses the dynamic behaviour to estimate positions of known objects and compare them to detected objects [6]. This project uses the motion model, which will be described in more detail in the following sections.

### 3.3.1 Object Detection

The most recent state of the art detection architecture published in the CVPR 2016 is discussed in the paper (YOLO) You Only Look Once: Unified, Real-Time Object Detection [9]. This is a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance. This unified architecture is extremely fast. the base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is less likely to predict false positives on background. Finally, YOLO learns very general representations of objects. It outperforms other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains like artwork. For our detection and localization stage we used an updated version of YOLO which is the YOLOV3 [10].

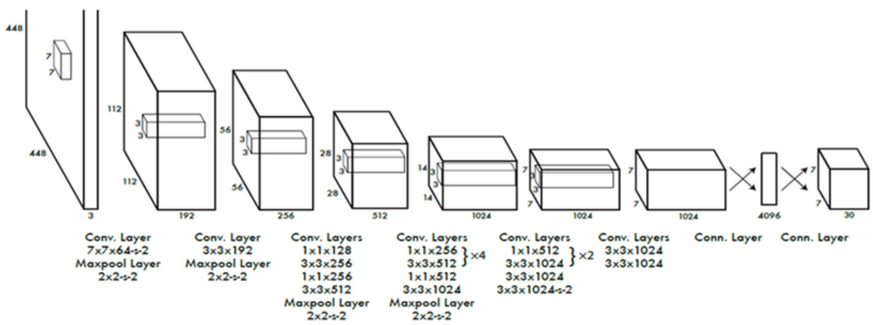


Figure 3.2: Architecture of YOLO

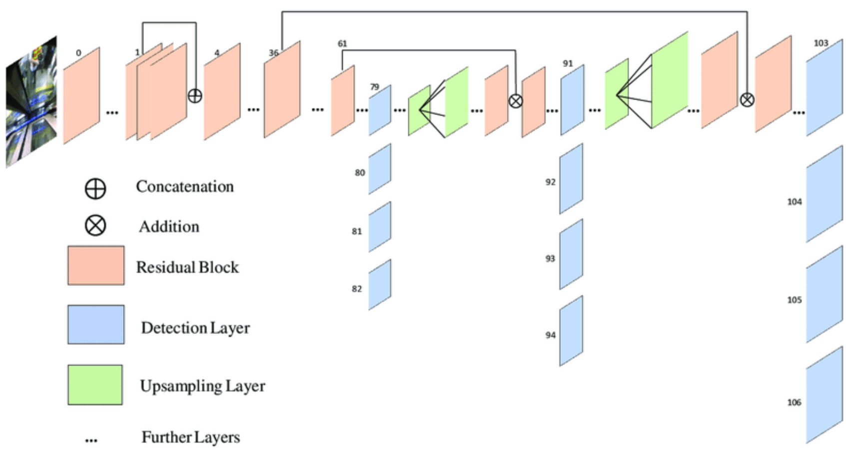


Figure 3.3: Architecture of YOLOV3

### 3.3.2 Motion Prediction

In the motion prediction stage, the aim is to estimate the new positions of the tracked objects from the previous frames. The predictions usually take the form of the object centroids or bounding boxes, and are used in the following steps to match and assign new detections to tracked objects. One common algorithm for this is the kalman filter [4]. The kalman filter uses a linear dynamical system to model the motion of the objects [5], which is used to estimate the centroids of all the tracked objects.

### 3.3.3 Affinity

In the affinity stage a cost or affinity matrix is created by calculating a score between each of the predicted and detected object positions. This score indicates how similar or how low the distance is between each pair of prediction and detections. The metric used to calculate the scores will depend on the specific implementation, for this project the metrics used are distance and IoU [4]. IoU, or intersection over union, is a metric for how big the overlap between two rectangles is and is used to determine how closely two bounding boxes match each other. The cost calculation is a ratio between the overlap and the union of the two bounding boxes.

### 3.3.4 Association

In the association stage the aim is to assign or match detected objects from the current frame to the tracked objects (tracks), and if necessary create new tracks for new objects, or remove tracks from objects no longer in the frame. The assignment is done using the cost matrix from the affinity stage, with the goal of matching detection/track pairs with the lowest costs. This can be solved using assignment problem algorithms such as the hungarian algorithm [8], which is an efficient algorithm for minimising the total cost of all pairs. To handle the birth and death of tracks, a detection is classified as a new track if it is not paired with an existing track, or if the cost of a pair is deemed too high. If a track is not paired with a detected object for a set amount of frames, it is deemed to have left the scene, and the track is removed.

## 3.4 Quantization

The multiple object tracking provides the track of each categorized object through out the feed. The tracks for each category are then counted as data instances of each category. This provides us with the final categorical quantization.

# Chapter 4

# Project Management & Financing

## 4.1 Work Breakdown Structure

The activities listed below were conducted on a bi-weekly basis :

- Data Acquisition & Preparation
- Model Configuration & Training
- Spill Video Validation
- Real Time Validation
- Analysing the detection and tracking activators
- Reviewing the pipeline to achieve better performance

## 4.2 Process/Activity wise Time Distribution

This is how the time of each agile sprint was distributed in between the backlog activities of each sprint:

- Data Acquisition & Preparation - 4 working days
- Model Configuration & Training - 1 working day
- Spill Video Validation - 2 working days
- Real Time Validation - 1 working day
- Analysing the detection and tracking activators - 1 working day
- Reviewing the pipeline to achieve better performance - 1 working day

## 4.3 Gantt Chart

| Activity | Week 1 | Week 2 | Week 3 | Week 4 | Week 5 | Week 6 | Week 7 | Week 8 |
|---|---|---|---|---|---|---|---|---|
| Data Acquisition & Preparation(Sprint 1) | 4/5days | | | | | | | |
| Model Configuration & Training(Sprint 1) | 1/5days | | | | | | | |
| Spill Video Validation(Sprint 1) | | 2/5days | | | | | | |
| Real Time Validation(Sprint 1) | | 1/5days | | | | | | |
| Activator Analysis(Sprint 1) | | 1/5days | | | | | | |
| Pipeline Review(Sprint 1) | | 1/5days | | | | | | |
| Data Acquisition & Preparation(Sprint 2) | | | 4/5days | | | | | |
| Model Configuration & Training(Sprint 2) | | | 1/5days | | | | | |
| Spill Video Validation(Sprint 2) | | | | 2/5days | | | | |
| Real Time Validation(Sprint 2) | | | | 1/5days | | | | |
| Activator Analysis(Sprint 2) | | | | 1/5days | | | | |
| Pipeline Review(Sprint 2) | | | | 1/5days | | | | |
| Data Acquisition & Preparation(Sprint 3) | | | | | 4/5days | | | |
| Model Configuration & Training(Sprint 3) | | | | | 1/5days | | | |
| Spill Video Validation(Sprint 3) | | | | | | 2/5days | | |
| Real Time Validation(Sprint 3) | | | | | | 1/5days | | |
| Activator Analysis(Sprint 3) | | | | | | 1/5days | | |
| Pipeline Review(Sprint 3) | | | | | | 1/5days | | |
| Data Acquisition & Preparation(Sprint 4) | | | | | | | 4/5days | |
| Model Configuration & Training(Sprint 4) | | | | | | | 1/5days | |
| Spill Video Validation(Sprint 4) | | | | | | | | 2/5days |
| Real Time Validation(Sprint 4) | | | | | | | | 1/5days |
| Activator Analysis(Sprint 4) | | | | | | | | 1/5days |
| Pipeline Review(Sprint 4) | | | | | | | | 1/5days |

Table 4.1: Gantt Chart

## 4.4 Process/Activity wise Resource Allocation

| Activity | Hardware | Software | Programming Language | Human |
|---|---|---|---|---|
| Data Acquisition | Surveillance Camera<br>The computer at XYZ Industries Limited server room<br>My office computer at Animo.AI | Google Drive | | Md Zafore Sadek<br>Rukon Uddin |
| Data Preparation | My office computer at Animo.AI | LabelImg | Python | d Zafore Sadek<br>Rukon Uddin |
| Model Configuration & Training | My office computer at Animo.AI<br>Google Colab Pro GPU | Google Drive<br>Google Colab Pro | Python | Md Zafore Sadek<br>Rukon Uddin |
| Spill Video Validation | My office computer at Animo.AI | Visual Studio Code<br>Pycharm<br>Google Drive | Python | Md Zafore Sadek<br>Rukon Uddin |
| Real Time Validation | The computer at XYZ Industries Limited server room<br>Nvidia Geforce RTX 2080ti GPU | Visual Studio Code<br>Google Drive<br>Pycharm<br>Teamviewer | Python | Md Zafore Sadek<br>Rukon Uddin |

Table 4.2: Process/Activity wise resource allocation

## 4.5 Estimated Costing

| Paid Hardware (BDT) | Paid Software (BDT) | Paid HR (BDT) | Total Amount (BDT) |
|---|---|---|---|
| 20000/= | 20000/= | 30000/= | 70000/= |

Table 4.3: Approximate estimation of costing

# Chapter 5

# Body of the Project

## 5.1 Requirement Analysis

### 5.1.1 Functional

- Function : Categorical Quantization.
- Input : Real time feed of objects from surveillance camera.
- Process : Detect, recognize and quantize objects categorically and send. count information to ERP system via API.
- Output : All objects will be categorically accounted for in real time.
- Precondition : Detector and recognizer must be pretrained.
- Postcondition : Real time categorical count of objects will be available in the ERP System.

### 5.1.2 Non-Functional

- Industry Grade Performance.
- Efficient GPU Utilization.
- Data Security & Integrity Assurance.
- Modular& Maintainable Solution.

## 5.2 System Analysis

| | Current | Proposed |
|---|---|---|
| Process | Categorical Quantization | Categorical Quantization |
| Human | Workers count the objects of each catogery manually. | No need for workers to count the objects of each catogery manually. |
| Non computing hardware | Pen, paper and registry books are used to record and store the count information. | Surveillence camera to capture feed from the conveyor belt. |
| Computing hardware | XYZ Industries Limited computers are used by operators to record the registry book data into the ERP System. | -XYZ Industries Limited computer present in the server room is used to store and upload necessary data to google drive.<br>-Animo computer used to download necessary data, prepare data, train model on colab, download model, run spill video tests using vs-code/pycharm and upload model to drive.<br>-Animo computer is used to access control of XYZ Industries Limited computer via teamviewer. Then model is downloaded on XYZ Industries Limited computer and deployed via pycharm on a GPU accelerated runtime environment to conduct real time validation of the feed.<br>-A RTX 2080ti is used for gpu accelaration on the XYZ Industries Limited computer. |
| Software | ERP System is used to monitor all data that is generated from the counting. | -Google Drive is used to store necessary data and make it accessible to Animo and XYZ Industries Limited.<br>-Colab is used to train the model and download the model.<br>-Teamviewer is used to access XYZ Industries Limited CPU from Animo.<br>-Model is downloaded from Google Drive at XYZ Industries Limited and Animo CPU.<br>-Pycharm and Vs-code is used for real time and spill video validation.<br>-A custom python module is used to send the real time validation data to the ERP System's SQL database. |
| Database | SQL database is used to store ERP System's information on the local sever. | SQL database is used to store ERP System's information on the local sever. |
| Network & Communication | Independent wired networking is used to maintain an interconnective network between the local server and computers at the industry and the computers at the head offices. | -Independent wired networking is used to maintain an interconnective network between the local server and computers at the industry and the computers at the head offices.<br>-Extra ISP provided networking is needed for XYZ Industries Limited server room cpu to share data with Animo. |

Table 5.1: Analysis of current and proposed utilization of the six elements

## 5.3 System Design

### 5.3.1 Current System

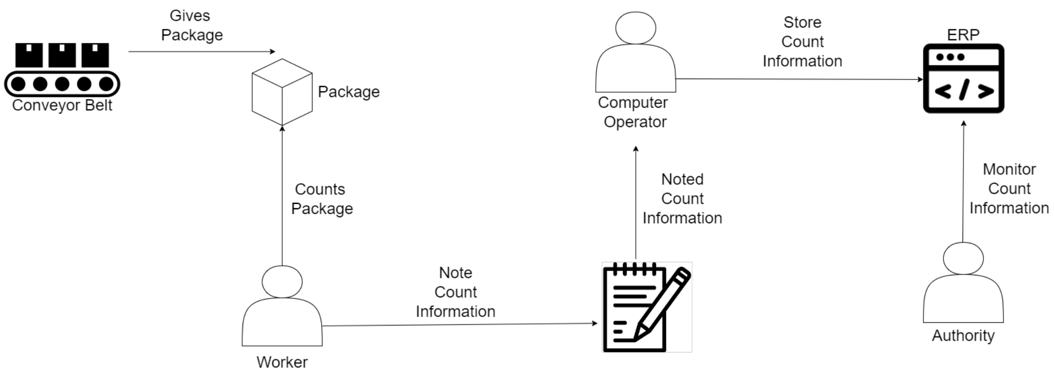


Figure 5.1: Current System

### 5.3.2 Proposed System

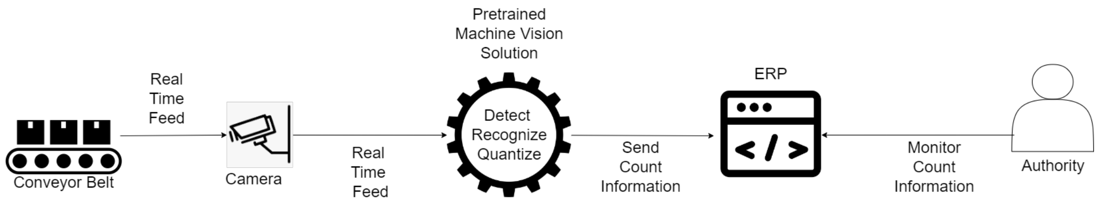


Figure 5.2: Proposed System

## 5.4 Work Description

In the first and second sprint we worked with the following 21 categories:

- Energy plus classic 60gm 24pk
- Energy plus tiffin pouch 24gm 120pk
- Energy plus tiffin chain 24gm 60pk
- Energy plus standard 80gm 24pkt
- Energy plus 240gm 6pkt
- Chokito white chocolate 360gm 6jar
- Chokito chocolate 360gm 6 jar
- Chokito white mango 360gm 6jar
- Chokito white orange 360gm 6jar
- Knock wafer strawberry 25gm 72pkt
- Knock wafer milk 25gm 72pkt
- Knock wafer chocolate 25gm 72pkt
- Nutty Standard 80gm 24pkt
- Nutty Family 250gm 8pkt
- Nutcream 100gm 24pkt
- Milkplus Standard 65gm 24pkt
- Olino Milk wafer 15gm 72pkt
- Olino Chocolate 15gm 72pkt
- First Choice Standard 100gm 24pkt
- Wayfun Milk Cream 20gm 72pkt
- First Choice Family 240gm 8pkt

### 5.4.1 Sprint 1

#### Week 1

Day 1

- Received feed from XYZ Industries Limited
- Split the feed into training and validation units.
- Spilt the validation unit into multiple spill videos.
- Each spill video is gone through manually to confirm ground truth on the categorical count the objects.

Day 2

- Annotated 4000 objects

Day 3

- Annotated 4000 objects

Day 4

- Annotated 3525 objects

Day 5

- Model Configuration
- Training

#### Week 2

Day 1

- Validated trained model on four spill videos with each video containing around 2 hours of feed.

Day 2

- Validated trained model on three spill videos with each video containing around 2 hours of feed.

Day 3

- Real time validation performed using RTSP[11] link from the mounted camera.

Day 4

- Analysis of validation results and the activation factors that contributed those results.

Day 5

- Constructing suggestions to improve performance.

### 5.4.2 Sprint 2

#### Week 3

Day 1

- Received new feed from XYZ Industries Limited
- Split the feed into training and validation units.
- Spilt the validation unit into multiple spill videos.
- Each spill video is gone through manually to confirm ground truth on the categorical count the objects.

Day 2

- Annotated 6500 new objects

Day 3

- Annotated 6500 new objects

Day 4

- Annotated 6000 new objects

Day 5

- Model Configuration
- Training

#### Week 4

Day 1

- Validated trained model on four spill videos with each video containing around 2 hours of feed.

Day 2

- Validated trained model on three spill videos with each video containing around 2 hours of feed.

Day 3

- Real time validation performed using RTSP[11] link from the mounted camera.

Day 4

- Analysis of validation results and the activation factors that contributed those results.

Day 5

- Constructing suggestions to improve performance.

Due to poor performance in the first and second sprint we scaled down on the categories to better understand the reasons behind such performance. Thus, In the third and fourth sprint we worked with the following 3 categories only:

- Energy plus classic 60gm 24pk
- Energy plus tiffin pouch 24gm 120pk
- Energy plus tiffin chain 24gm 60pk

### 5.4.3 Sprint 3

#### Week 5

Day 1

- Received new feed from XYZ Industries Limited
- Split the feed into training and validation units.
- Spilt the validation unit into multiple spill videos.
- Each spill video is gone through manually to confirm ground truth on the categorical count the objects.

Day 2

- Annotated 4000 objects

Day 3

- Annotated 4000 objects

Day 4

- Annotated 3602 objects

Day 5

- Model Configuration
- Training

#### Week 6

Day 1

- Validated trained model on four spill videos with each video containing around 2 hours of feed.

Day 2

- Validated trained model on three spill videos with each video containing around 2 hours of feed.

Day 3

- Real time validation performed using RTSP[11] link from the mounted camera.

Day 4

- Analysis of validation results and the activation factors that contributed those results.

Day 5

- Constructing suggestions to improve performance.

### 5.4.4 Sprint 4

#### Week 7

Day 1

- Received new feed with better illumination from XYZ Industries Limited
- Split the feed into training and validation units.
- Spilt the validation unit into multiple spill videos.
- Each spill video is gone through manually to confirm ground truth on the categorical count the objects.

Day 2

- Annotated 6000 new objects

Day 3

- Annotated 6000 new objects

Day 4

- Annotated 6119 new objects

Day 5

- Model Configuration
- Training

### Week 8

Day 1

- Validated trained model on four spill videos with each video containing around 2 hours of feed.

Day 2

- Validated trained model on three spill videos with each video containing around 2 hours of feed.

Day 3

- Real time validation performed using RTSP[11] link from the mounted camera.

Day 4

- Analysis of validation results and the activation factors that contributed those results.

Day 5

- Constructing suggestions to improve performance.

This is a log of the eight week research and development conducted by me during the internship. In the following section, there will be a detailed discussion on the results of each sprint.

## 5.5 Output Demonstration

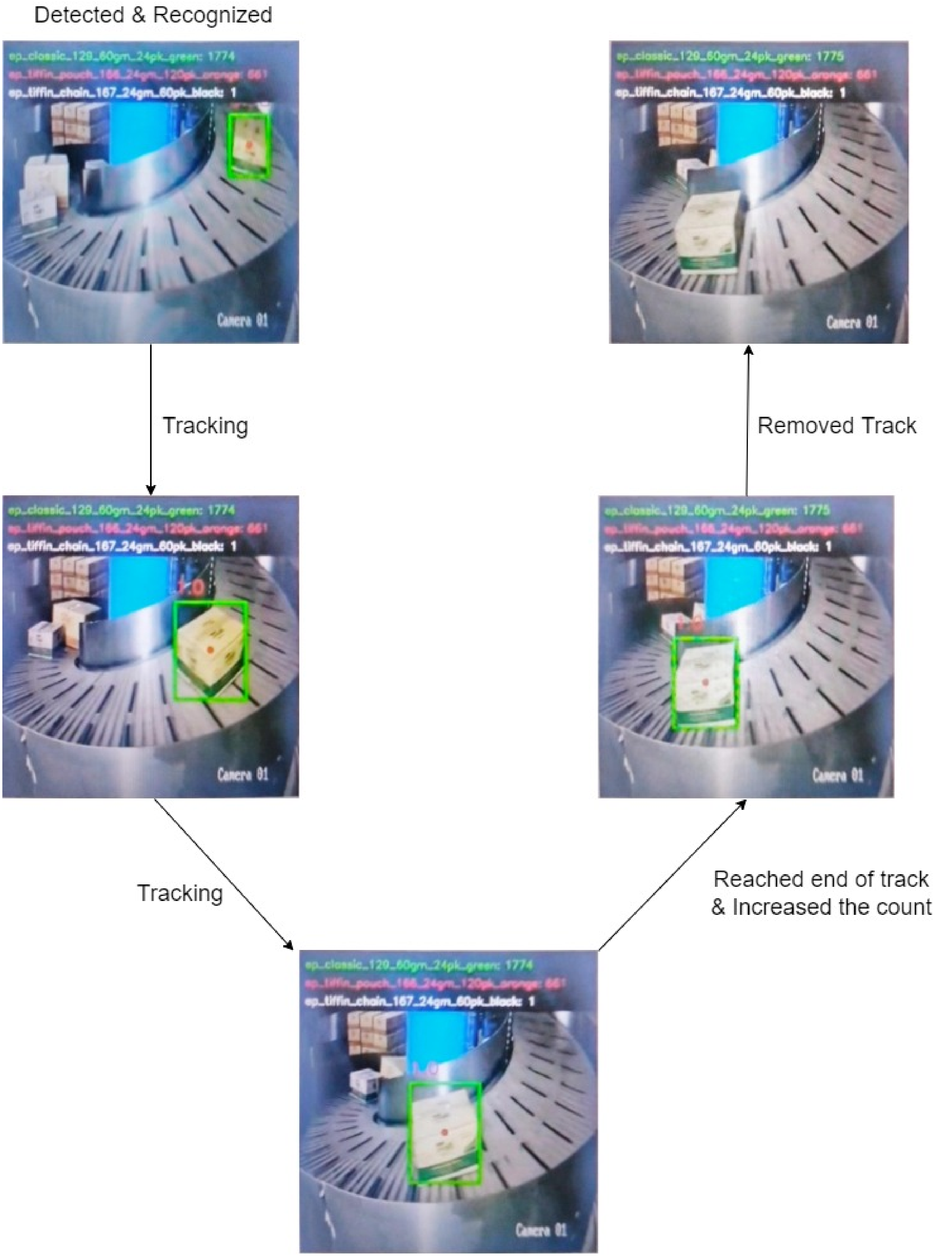


Figure 5.3: Output Demonstration

# Chapter 6

# Results & Analysis

## 6.1 Sprint 1

### 6.1.1 Data Analysis

| No. | Product Name | Total Datapoint |
|---|---|---|
| 1 | Energy plus classic 60gm 24pk (green) | 2828 |
| 2 | Energy plus tiffin pouch 24gm 120pk (orange) | 1246 |
| 3 | Energy plus tiffin chain 24gm 60pk (black) | 112 |
| 4 | Energy plus standard 80gm 24pkt | 376 |
| 5 | Energy plus 240gm 6pkt | 452 |
| 6 | Chokito white chocolate 360gm 6jar | 515 |
| 7 | Chokito chocolate 360gm 6 jar | 756 |
| 8 | Chokito white mango 360gm 6jar | 656 |
| 9 | Chokito white orange 360gm 6jar | 533 |
| 10 | Knock wafer strawberry 25gm 72pkt | 237 |
| 11 | Knock wafer milk 25gm 72pkt | 292 |
| 12 | Knock wafer chocolate 25gm 72pkt | 189 |
| 13 | Nutty Standard 80gm 24pkt | 341 |
| 14 | Nutty Family 250gm 8pkt | 345 |
| 15 | Nut&cream 100gm 24pkt | 348 |
| 16 | Milkplus Standard 65gm 24pktz | 326 |
| 17 | Olino Milk wafer 15gm 72pkt | 293 |
| 18 | Olino Chocolate 15gm 72pkt | 507 |
| 19 | First Choice Standard 100gm 24pkt | 296 |
| 20 | Wayfun Milk Cream 20gm 72pkt | 459 |
| 21 | First Choice Family 240gm 8pkt | 418 |

Table 6.1: Number of train data points in each category.

### 6.1.2 Accuracy

The average spill video validation accuracy for the first sprint was 26.21% and the real time validation accuracy was 18.9%.

### 6.1.3 Sprint Analysis

After analysis it was clear that the imbalance of data points led to the poor performance. Thus, XYZ Industries Limited was to provide more feed so that the data points could be more balance in the second sprint. Moreover, it should be mentioned that the positional orientation of the test objects in the real time validation was very different from the positional orientation of the train objects in the train feed. This led to the huge difference between the spill video validation accuracy and the real time validation accuracy.

## 6.2 Sprint 2

### 6.2.1 Data Analysis

| No. | Product Name | Total Datapoint |
|---|---|---|
| 1 | Energy plus classic 60gm 24pk (green) | 2828 |
| 2 | Energy plus tiffin pouch 24gm 120pk (orange) | 1246 |
| 3 | Energy plus tiffin chain 24gm 60pk (black) | 1112 |
| 4 | Energy plus standard 80gm 24pkt | 1376 |
| 5 | Energy plus 240gm 6pkt | 1452 |
| 6 | Chokito white chocolate 360gm 6jar | 1515 |
| 7 | Chokito chocolate 360gm 6 jar | 1756 |
| 8 | Chokito white mango 360gm 6jar | 1656 |
| 9 | Chokito white orange 360gm 6jar | 1533 |
| 10 | Knock wafer strawberry 25gm 72pkt | 1237 |
| 11 | Knock wafer milk 25gm 72pkt | 1292 |
| 12 | Knock wafer chocolate 25gm 72pkt | 1189 |
| 13 | Nutty Standard 80gm 24pkt | 1341 |
| 14 | Nutty Family 250gm 8pkt | 1345 |
| 15 | Nut&cream 100gm 24pkt | 1348 |
| 16 | Milkplus Standard 65gm 24pktz | 1326 |
| 17 | Olino Milk wafer 15gm 72pkt | 1293 |
| 18 | Olino Chocolate 15gm 72pkt | 1507 |
| 19 | First Choice Standard 100gm 24pkt | 1296 |
| 20 | Wayfun Milk Cream 20gm 72pkt | 1459 |
| 21 | First Choice Family 240gm 8pkt | 1418 |

Table 6.2: Number of train data points in each category.

In the second sprint, 1000 more data points were added in the 19 classes that did not have at least 1000 data points.

### 6.2.2 Accuracy

The average spill video validation accuracy for the first sprint was 88.63% and the real time validation accuracy was 80.88%.

### 6.2.3 Sprint Analysis

A huge improvement in performance can be observed due to balancing the number of data points. However, this cannot be considered industry grade performance. Thus, the plan for the next sprint is to scale down on the categories to understand how each category is loosing accuracy.

## 6.3 Sprint 3

### 6.3.1 Data Analysis

| No. | Product name | Total datapoints |
|---|---|---|
| 1 | ep_classic_129_60gm_24pk (green) | 4365 |
| 2 | ep_tiffin_pouch_166_24gm_120pk (orange) | 2882 |
| 3 | ep_tiffin_chain_167_24gm_60pk (black) | 4355 |

Table 6.3: Number of train data points in each category.

### 6.3.2 Accuracy

| | | Product Variant | | | |
|---|---|---|---|---|---|
| Spell | Energy plus 60 gm Accuracy (%) | Energy plus 24gm (120 pk) Accuracy (%) | Energy plus 24gm (60 pk) Accuracy (%) | Remarks | Accuracy (%) |
| Spill 1 | 85 | 95 | 80 | Miss Detection | 88 |
| Spill 2 | 93 | 95 | 90 | Miss Detection | 93 |
| Spill 3 | 96 | 90 | 98 | Miss Detection | 92 |
| Spill 4 | 92 | 93 | 90 | Miss Detection | 95 |
| Spill 5 | 82 | 86 | 90 | Miss Detection | 90 |
| Spill 6 | 96 | 92 | 95 | Miss Detection | 95 |
| Spill 7 | 98 | 93 | 95 | Miss Detection | 94 |

Table 6.4: In-depth accuracy analysis for spill video validation.

The average spill video validation accuracy was 92.42% and the real time validation accuracy was 88.77%.

### 6.3.3 Sprint Analysis

After extensive analysis it was clear that there was a lack of illumination that led to the accuracy loss in each category. So we suggested increasing the illumination at the place where the camera was mounted and acquiring new data.

## 6.4 Sprint 4

### 6.4.1 Data Analysis

| No. | Product name | Total datapoints |
|---|---|---|
| 1 | ep_classic_129_60gm_24pk (green) | 8209 |
| 2 | ep_tiffin_pouch_166_24gm_120pk (orange) | 3843 |
| 3 | ep_tiffin_chain_167_24gm_60pk (black) | 6067 |

Table 6.5: Number of train data points in each category (with improved illumination).

### 6.4.2 Accuracy

| | | Product Variant | | | |
|---|---|---|---|---|---|
| Spell | Energy plus 60 gm Accuracy (%) | Energy plus 24gm (120 pk) Accuracy (%) | Energy plus 24gm (60 pk) Accuracy (%) | Remarks | Accuracy (%) |
| Spell 1 | 100 | 100 | 100 | | 100 |
| Spell 2 | 98.75 | 99 | 97.91 | Power outage | 98.5 |
| Spell 3 | 99.58 | 100 | 99.48 | Double bounding box | 99.68 |
| Spell 4 | 99.58 | 100 | 98.43 | Double bounding box | 99.72 |
| Spell 5 | 99.58 | 100 | 99.48 | Double bounding box | 99.68 |
| Spell 6 | 99.58 | 100 | 98.43 | Double bounding box | 99.72 |
| Spell 7 | 100 | 100 | 100 | | 100 |

Table 6.6: In-depth accuracy analysis for spill video validation (with improved illumination).

The average spill video validation accuracy was 99.61% and the real time validation accuracy was 96.8%.

### 6.4.3 Sprint Analysis

We conclude our project proceedings with this sprint. The outstanding performance in this sprint enforces proof of the concept that the activity of categorically counting product packages with an automated computer vision approach is plausible and can deliver industry grade performance.

# Chapter 7

# Lesson Learned

## 7.1 Problems Faced During this Period

- Data acquisition was a time consuming task due to the data being stored on XYZ Industries Limited local server.
- As we are following a supervised learning approach annotating ground truth for all objects in every sprint was a tedious task.
- Due to XYZ Industries Limited using a local server the only way to deploy the model for real time validation was to use team viewer and use their computer to deploy the model on their server.
- Data imbalance in the first sprint.
- low resolution camera and low fps feed.
- During a power outage, everything else except the camera had backup power.
- As this solution will enable the authorities to cut down on employment, the workers are reluctant in their co-operation during real time validation phases. They would try to alter the positional orientation of the objects; sometimes they would put in object of categories that we have not trained the model on, sometimes they would pile up objects and then release them altogether onto the conveyor; a scenario that the model was not familiar with.

## 7.2 Solution of those Problems

- Transition to a cloud server would increase accessibility and improve data acquisition and model deployment issues.
- Further research is needed with agenda of automating identification of the region of interest for each object so that an unsupervised learning approach can be undertaken to cluster the objects of each category. This may solve the tedious task of manual annotations.
- Data imbalance can be addressed easily if the transition to cloud makes the feed easily accessible.
- A representative from the authorities at XYZ Industries Limited should be briefed about the scenarios the model is familiar with before real time validation sessions.
- A camera that has high frames per second, and a very good image resolution is required. Some good cameras are Avalonix , lilin UFG1122E, Nelly's security camera.

# Chapter 8

# Future Work & Conclusion

## 8.1 Future Works

For future work there are a few main areas to look at. First of all, as good as the counting results are, there are still improvements to be made for the MOT algorithm to tackle identity switches and fragmented tracks. This is especially significant for inspection methods that rely on tracks staying intact.

A possible area to look into in this regard is the use of recurrent neural networks for the affinity and association stages, as proposed in [7]. This could potentially lower the number of identity switches and fragmented tracks, resulting in better performance.

Some other prospective research that can be applicable for this project include studying how the region of interest for each object can identified automatically and working towards intelligent modules for product and package inspection on conveyor belts.

## 8.2 Conclusion

I would like to mention that this project has acquired the proof of concept accreditation from XYZ Industries Limited authorities and this project will be continued to be deployed on full scale. For full scale deployment we will be performing machine vision based categorical quantization on 63 categories spread over 100 conveyor belts.

Concluding this report by once again acknowledging this enlightening experience for what it was and expressing my gratitude towards my internal and external supervisors for their guidance throughout this journey.

# Bibliography


[1] Mehmet Baygin and Mehmet Karakose. A new image stitching approach for resolution enhancement in camera arrays. In *2015 9th International Conference on Electrical and Electronics Engineering (ELECO)*, pages 1186–1190. IEEE, 2015.

[2] Mehmet Baygin, Mehmet Karakose, Alisan Sarimaden, and Erhan Akin. An image processing based object counting approach for machine vision application. *arXiv preprint arXiv:1802.05911*, 2018.

[3] Vebjørn Bjørlo-Larsen. Vision based real-time fish counting, inspection and classification using deep learning. Master's thesis, NTNU, 2021.

[4] Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, and Francisco Herrera. Deep learning in video multi-object tracking: A survey. *Neurocomputing*, 381:61–88, 2020.

[5] Felix Govaers. *Introduction and Implementations of the Kalman Filter*. BoD–Books on Demand, 2019.

[6] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple object tracking: A literature review. *Artificial Intelligence*, 293:103448, 2021.

[7] Anton Milan, S Hamid Rezatofighi, Anthony Dick, Ian Reid, and Konrad Schindler. Online multi-target tracking using recurrent neural networks. In *Thirty-First AAAI conference on artificial intelligence*, 2017.

[8] G Ayorkor Mills-Tettey, Anthony Stentz, and M Bernardine Dias. The dynamic hungarian algorithm for the assignment problem with changing costs. *Robotics Institute, Pittsburgh, PA, Tech. Rep. CMU-RI-TR-07-27*, 2007.

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[10] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[11] Henning Schulzrinne, Anup Rao, and Robert Lanphier. Real time streaming protocol (rtsp). 1998.