Smart Vision Technology Quality Control (Robotics)

Team Name: sethisubhanshu

Date of Submission: 20/10/2024

Sidh Dhiman (JIIT NOIDA)

Subhanshu Sethi (DTU)

Flipkart | GRiD 6.0

EXECUTIVE SUMMARY

Extraction of Details from Packaging Material

- We utilized the **Qwen2-VL Vision Language Model** to extract key details from packaging materials, such as (**Brand Name, Product Type, Expiry Date, Manufacturing Date, Quantity)**
- On our customized test data for **Brand Name** and **Product Type**, we achieved an impressive **F1 score of 0.86**.

Processed Output

Qwen Output: Brand Name: Nestle
Product Type: Dark Chocolate Coated Wafer
Expiry Date: July 25
Manufacturing Date: August 24
Quantity: 150g

Freshness and Shelf Life Prediction

- For predicting freshness and shelf life of perishable goods, we applied a **lighter version of the Qwen2-VL model**. The model outputs the remaining **shelf life** of fruits and vegetables, helping in accurate freshness assessment.

Freshness Detection Output

The fruit in the image is a banana. It appears to be worse in freshness. It has visible signs of spoilage such as discoloration and potential mold, indicating that it may not be suitable for consumption.

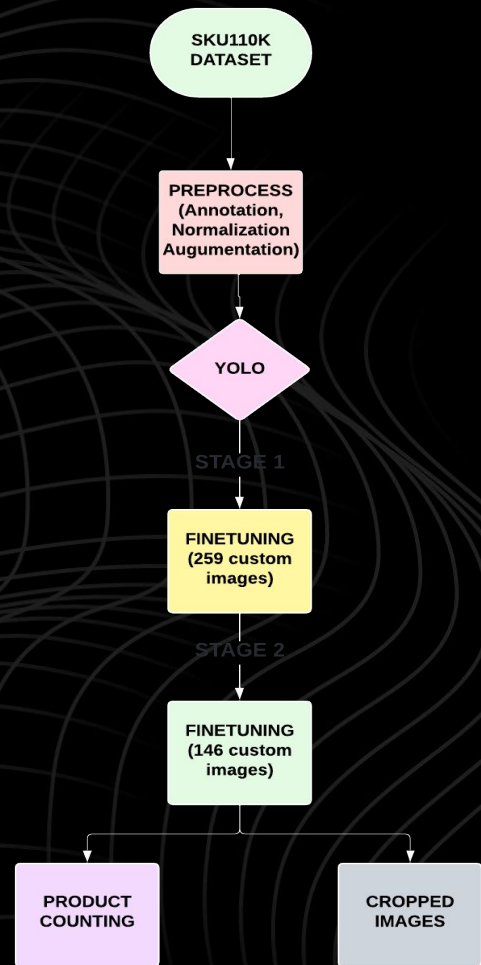Estimated Shelf Life for Banana with Worse: less than 2 days

DATA COLLECTION AND CURATION

YOLO Model

- **SKU110K Dataset**: Used as a base for training the **YOLOv5/v8 model**, leveraging its extensive grocery product data.
- **Custom Dataset**:
  - **Stage 1**: Fine-tuned the YOLO model with a **custom dataset of 259 annotated images**, tailored to our specific task.
  - **Stage 2**: Further fine-tuned the model with an additional **146 custom images** to enhance detection performance.
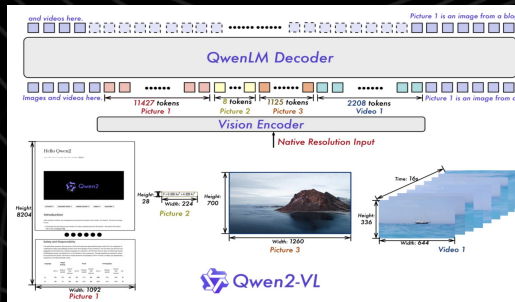
Qwen2-VL Model

- **User Message Generation (Freshness Classification)**:
  - **Classification Dataset**: Utilized images of fresh and rotten fruits/vegetables for the freshness classification task.
  - **Message Generation**: Created custom messages based on freshness classification to provide tailored outputs.
- **User Message Generation (Product Description)**:
  - **Custom Dataset**: Collected datasets from nearby stores, focusing on relevant product types.
  - **Message Generation**: Developed custom messages for extracting useful product information, such as brand name, expiry date, and quantity.

SKU110K DATASET

PREPROCESS (Annotation, Normalization Augumentation)

YOLO

STAGE 1

FINETUNING (259 custom images)

STAGE 2

FINETUNING (146 custom images)

PRODUCT COUNTING

CROPPED IMAGES

TECHNICAL IMPLEMENTATION

Why Qwen2-VL?

1. High Accuracy
2. Open Source
3. Vision-Language Alignment
4. Pretrained on OCR & VQA Datasets

DocVQA$_{test}$	91.6	90.8	–	**94.5**
InfoVQA$_{test}$	74.8	–	–	**76.5**
ChartQA$_{test}$	**83.3**	–	–	83.0
TextVQA$_{val}$	77.4	80.1	–	**84.3**
OCRBench	794	**852**	785	845

Fine Tuning approach(Product Descriptor)
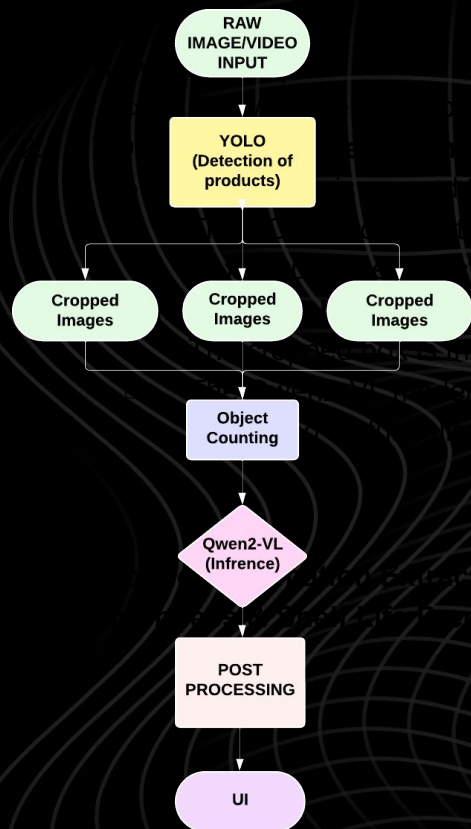
❖ **Qwen2-VL 7B**

- **First Fine-Tuning Stage:** Fine-tuned the model on a custom dataset of **90** images for **25** epochs.
- **Second Fine-Tuning Stage:** Conducted an additional **20** epochs of fine-tuning with **60** images, incorporating varied user messages to handle edge cases.

Fine Tuning approach(Freshness)

❖ **Qwen2-VL 2B**

- **Fine-Tuning:** Fine-tuned the model for **45** epochs with dataset of **550** images, generating custom user messages tailored to the requirements of shelf-life prediction.

PIPELINE(PRODUCT DESCRIPTOR)

```
RAW IMAGE/VIDEO INPUT
        ↓
YOLO (Detection of products)
        ↓
Cropped Images    Cropped Images    Cropped Images
        ↓
Object Counting
        ↓
Qwen2-VL (Infrence)
        ↓
POST PROCESSING
        ↓
UI
```
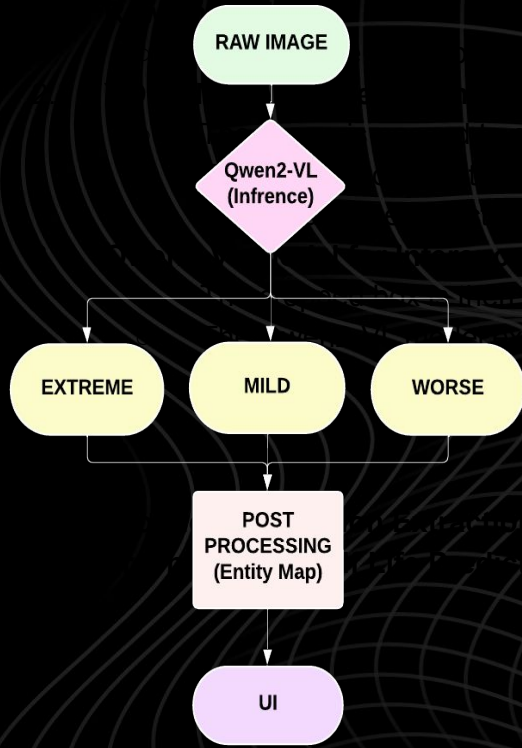
Approach:

- **Raw Image Input:**
  - A raw image of the packaging material is captured and processed.

- **YOLO Model for Detection:**
  - The image is passed to our custom YOLO model for object detection.
  - The YOLO model detects relevant regions on the packaging (e.g., brand name, expiry date) and provides a cropped bounding box around the detected area.

- **Qwen2-VL Model for Inference:**
  - The cropped box is then passed to the Qwen2-VL model for further inference.
  - The Qwen2-VL model extracts detailed information such as the brand name, expiry date, and other key attributes, based on visual and textual alignment.

Stage-1	Stage-2
Base Model: Qwen2-Vl-7B-Instruct	Base Model: Qwen2-Vl-7B-Instruct
Learning Rate:5e-4	Learning Rate:4e-4
FIneTuning Methods: SFTTrainer, QLora (4 bit), PEFT	FIneTuning Methods: SFTTrainer, QLora (4 bit), PEFT
PROMPT USED: "Describe the product in given image"	PROMPT USED: "Identify the brand name, product type, expiry date, manufacturing date, quantity only."
Dataset SIze: 90 images	Dataset SIze: 90 images
Hyperparameters used: Batch_size: 4, Epochs: 25, LR_scheduler: Cosine	Hyperparameters used: Batch_size: 2, Epochs: 20, LR_scheduler: Constant

PIPELINE(FRESHNESS)

RAW IMAGE

↓

Qwen2-VL (Infrence)

↓

EXTREME **MILD** **WORSE**

↓

POST PROCESSING (Entity Map)

↓

UI

Approach:
- **Raw Image Input:**
  - The process begins with feeding raw images of fresh and rotten fruits/vegetables into the Qwen2-VL 2B model. These images are unprocessed and represent various stages of freshness.
- **Model Classification:**
  - The Qwen2-VL 2B model processes these images and classifies them into three categories:
    - Extreme
    - Mild
    - Worse
- **Entity Mapping:**
  - Once the model outputs these classifications, an entity map is used to translate the categories into shelf life of the produce. This mapping ensures consistency and relevance to the freshness assessment task.

Steps

Base Model: Qwen2-Vl-2B-Instruct

Learning Rate:5e-5

FIneTuning Methods: SFTTrainer, QLora (4 bit), PEFT

PROMPT USED: "Identify the fruit or vegetable in the image and strictly categorize its freshness as one of the following: 'extreme freshness,' 'mild freshness,' or 'worse."

Dataset SIze: 550 images

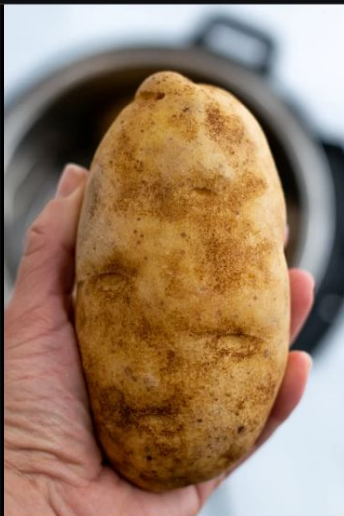**Hyperparameters used: Batch_size: 4, Epochs: 50, LR_scheduler: Constant**

RESULTS

Processed Output

Qwen Output: Brand: Amul, Product Type: Milk, Expiry Date: 10 December 2024, Manufacturing Date: 11 September 2024, Quantity: 450 mL

Upload Image

Freshness Detection Output

The fruit or vegetable in the image is a raw potato. It appears to be of moderate freshness. The color is uniform, the texture is firm, and there are no visible signs of spoilage.

Estimated Shelf Life for Potato with Mild freshness: around 10-12 days

Required Hardware:
Model Training: T4 GPU(Kaggle) x2
Model Inference:
1. Product Description(Qwen2-VL 7B)
   - NVIDIA L4 (18 GB VRAM)
2. Freshness Detector(Qwen2-VL 2B)
   - NVIDIA L4 (10 GB VRAM)

F1 Score (Brand Name and Product Type)
   - 0.86
YOLO inference time: 11ms
Qwen2-VL 7B inference time: 2.70 sec
Qwen2-VL 2B inference time: 1.90 sec

SCALIBLITY

FAST INFRENCE TIME

HIGH ACCURACY

HIGH COMPTIBLITY

CAN BE EXTENDED TO VARIOUS MORE TASKS

MINIMAL DATASET REQUIRED TO TRAIN

Flipkart
GRID
6.0

THANK YOU

FUTURE SCOPE

INTRODUCING OBJECT TRACKING TO TRACK OBJECT ID FOR FASTER REAL TIME INFERENCE

LARGER TRAINING WITH MORE DATASET TO INCREASE ACCURACY

DETECTING FRESHNESS OF BUNCH OF ITEMS

END TO END PRODUCTION IMPLEMENTAION

INVENTORY MANAGEMENT

USING MORE EFFICIENT INFERENCE TECHNIQUES