Visual Question Answering System

FINAL PROJECT

TIER 2

FRANCISCO RENTERIA RIOS

Visual Question Answering System

The problem

- The difficulty visually impaired individuals face when attempting to read the grocery labels, make them make mistakes when choosing store items.
- This project demonstrates a Visual Question Answering (VQA) system designed to help visually impaired users understand grocery labels and small print documents.
- It is important to support people with visual disabilities and make them more independent by inclusion using technology.

Visual Question Answering System

The Solution

- A **Visual Question Answering (VQA)** system takes an image and a question about that image, then generates an answer.
- For example, this system can be an app in a smart phone. The user ask a question, “Does this contain nuts?” They system then “sees” the label, understand the text, and gives the specific answer to the question, “Yes, the ingredient list includes almonds and peanuts”.
- This system empowers visually impaired users to shop, choose, and eat safely.

Visual Question Answering System

Technical Approach

- Pre-trained BLIP-2 model (Bootstrapped Language Image Pretraining). BLIP-2 has an excellent balance of performance, efficiency, and strong reasoning capability.
- Framework: PyTorch or TensorFlow for deep learning
- Platform: Google Colab
- EasyOCR reader

Visual Question Answering System

Dataset Plan

- Source: Public free dataset (VizWiz-VQA). It can be access through Hugging Face, Kaggle, or the official VizWiz website. (VizWiz dataset was created specifically for visually impaired people).
- Size: VizWiz contains over 31,000 image/question pairs.
- Labels: If a specific ingredient is in a product.
- Preparation: The dataset already contains labeled data. Although the data contains images with poor quality to represent a real world scenario.

Visual Question Answering System

System Diagram

- A simple pipeline flow:
- [Choose a Pre-trained VQA Model] → [Select an image] → [Ask a question related to the image] → [Feed the image and question into the model] → [Get the predicted answer] → [Evaluate accuracy]

- The application will download 10 sample images and make prediction based on predefined questions. It is true that VizWiz are not exclusevely grocery items labels, the images will be tested for their question/answer results. Initially, I plan to have static predictions, but a future plan would be a random lable image taken from a phone.

Visual Question Answering System

Success Metrics

- Average Normalized Levenshtein Similarity (ANLS): it measures how similar the predicted text is to the correct text.
- Answerability Prediction (this is a crucial metric drawn from the VizWiz dataset): The system must first decide if the question is even answerable from the image (e.g., the label is blurry, cut off, or doesn't contain the info).
- Optical Character Recognition (OCR) Metrics: The VQA's performance is capped by its ability to read.

Visual Question Answering System

Week-by-Week Plan

Week	Task	Milestone
10 (Oct 30)	Get dataset, setup environment	Dataset ready
11 (Nov 6)	Train of fine-tune model	Model working
12 (Nov 13)	Test and improve	Good accuracy
13 (Nov 20)	Create demo / video	Demo ready
14 (Nov 27)	Final testing / documentation	Everything done
15 (Dec 4)	Present project	Presentation day

Visual Question Answering System

Challenges & Backup Plans

- Poor Image Quality due to blur, the labeled can be cropped, poor lighting, or a finger can be covering the camera lens.
- Items in arbitrary poses and perspectives make difficult to locate and recognize the relevant text.
- Label reflection and glare. When reading the image, the glossy label material under store lighting makes the capture of the image look obscure.
- Labels and packaging use a variety of fonts, font sizes, colors, graphics, and background patterns.
- Training data must reflect the real-world conditions encounter by visually impaired conditions. And answers must be accurate.

Visual Question Answering System

Resources Needed

Resource	Options / Notes
Compute	Google Colab, VizWiz-VQA, Heidi
Frameworks	Pytorch
Estimated Cost	$0.00

GitHub Repository

https://github.com/francisco-renteria-rios/VQA_Assistance.git