Visual Question Answering

Sprint 1

By

Ajay Krishna Anand
Mingda Li
Saurabh Singh
Shangzhou Yin

Product Mission

- To help individuals with visual impairment through the use of Visual Question Answering technology.
- Achieved by - **analyzing a picture taken by the user and then returning an appropriate response based on the question asked** (taken either through a verbal or a text-based interface).
- Our goal is to adapt and improve previously made Open Source models for our specific purposes.

Q: Does this foundation have any sunscreen?

A: yes

Q: What is this?

A: 10 euros

Q: What color is this?

A: green

Q: Please can you tell me what this item is?

A: butternut squash red pepper soup

Q: Is it sunny outside?

A: yes

Q: Is this air conditioner on fan, dehumidifier, or air conditioning?

A: air conditioning

Q: What type of pills are these?

A: unsuitable image

Q: What type of soup is this?

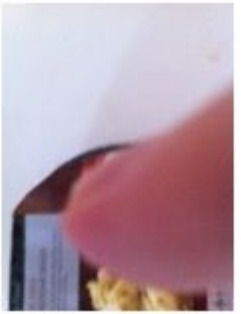
A: unsuitable image

Q: Who is this mail for?

A: unanswerable

Q: When is the expiration date?

A: unanswerable



Q: What is this?

A: unanswerable

Q: Can you please tell me what the oven temperature is set to?

A: unanswerable

Our Users and User Stories

We are focussing on building our product for **people with visual impairments.**

- As an individual with visual impairment, I want the product to tell me exactly what is in front of me so that I could prevent myself from colliding with it.
- As an individual with visual impairment, I want the product to describe me the details of an item, such as the shape, size and colour, so that I could interact with them.
- As an individual with visual impairment, I want the product to tell me where specific objects are located so that I could find them easily.

Literature Review

Title	Summary
Visual Question Answer Diversity - 2018 (https://aaai.org/ocs/index.php/HCOMP/HCOMP18/paper/viewFile/17936/16923)	This paper discusses the fact that a visual question can lead to a single answer or multiple answers. They propose a model that predicts the answer distribution that would be expected from a crowd for a given visual question.
VizWiz Grand Challenge: Answering Visual Questions from Blind People - 2018 (https://arxiv.org/abs/1802.08218)	The paper introduces a VQA dataset called Vizwiz and tests its utility.
Vision Skills Needed to Answer Visual Questions - 2020 (https://www.ischool.utexas.edu/~dannag/publications/2020_CSCW_VQA-Skills.pdf)	It identified the skill needed for assisting populations with visual impairments as well as a visual Turing test for the AI community.
Visual question answering: A survey of methods and datasets(https://www.sciencedirect.com/science/article/abs/pii/S1077314217300772)	It introduces the methods and datasets for VQA, and the future direction about it.
Visual Challenges in the Everyday Lives of Blind People-2012(https://www.cs.cmu.edu/~jbigham/pubs/pdfs/2013/visualchallenges.pdf)	The paper reports the findings of a large-scale study of the visual questions that blind people would like to have answered.

MVP

Input

- An image
- A free-form, open-ended, natural language question

Output

- Natural language answer

https://www.trolleytours.com/boston/acorn-street

e.g. Are there any people in front of me?
What's the name of the street?

Technologies to Evaluate

Visual Question answering could be broken down into four basic steps:

- Image Featurization - To process images usually using CNN
- Question Featurization - To convert natural language questions into a form that is processable.
- Joint Featurization Representation - to process both, features from images and text, at the same time.
- Answer Generation

Another equally important decision is the selection of the dataset used to train the model, as each dataset had certain inherent biases and difficulties in determining evaluation metrics.

Setup of Development Environment

Python + PyTorch

- PyTorch has a reputation for simplicity, ease of use, flexibility, efficient memory usage, and dynamic computational graphs. It also feels native, making coding more manageable and increasing processing speed.

- Plus all the recent papers and models related to VQA is developed in Pytorch. This led us to choose PyTorch!

Next Sprint Goals

- Set up development environment (Python + PyThorch).
- Evaluate and choose a dataset suitable for our purposes.
- Research Open Source Projects that could be adapted to build our product.

Thank You!