Dell Hackathon - Problem 2

Context based image & Text description

Problem Statement Description : Context based image description

- **Visually challenged students** do not have the privilege to view images. Describing the images in words will help them to understand and appreciate the images.
- Analyzing the image and describing the image content is quickly catching up, but in the case of students the images in textbooks are more context based and support the content around the image.
- **Ask** – Develop an intelligent system which will analyze the image and the context of the content around the image and describe the image accordingly.

Expected Outcome

- Read the contend from the file provided
- Analyze the image, content and build the context
- Describe the image based on the overall context built.
- English and multiple (around 2 to 3) Indian languages.

Tools & Technologies

- Python, OpenCV, Gen – AI (if required) or any other algorithms which fits the requirement (use only open-source algorithms/technologies)

Data Dependency

- No dependency, use any academic scanned textbooks for this process.

Assumptions/callouts

- Can use any pre-trained opensource models.
- Do not use any cloud or licensed solutions.
- Any scanned textbook can be used for training/building/testing/demo.

Context based image & Text description

Just image interpretation: **Man standing in the field**

Actual interpretation after combining text & image: **Lencho standing in between the field which is damaged due to the hail rain.**

Analyze the text around the image

Analyze the image

Summarize the image based on the text and image analyzed

Just image interpretation: **People cheering**

Actual interpretation after combining text & image: **Scientists cheering, clapping and jumping on successful launch of Chandrayaan 3**

Solution breakdown

Parts

Image Extraction
- Need to figure out how to approach this
- If the pdf is a text encoded pdf, we should be able to extract the image by simply parsing the pdf
- if not text encoded but a scanned type image pdf than need to see what to do

Text Extraction
- Fairly straightforward
- If text encoded pdf, just parse using (Py2PDF or similar)
- if scanned doc, than need to run ocr, tesseract or something similar

Context Generation Model
- Given the image and the text, use it to extract context
- Need to identify potential models to use
-

Indian Language Translation
- Need to figure out how to translate output from english context to indian language
- should be fairly simple, worst case just use google translate

Text to Speech
- Potential USP
- since the target audience cannot see, play text back to them
- if time constraint than use generic py, text2speech, else bark.ai for custom voice

UI
- Need to figure out
- Bootstrap + react/angular/vue

Potential Issues
- How to segment image from text ?