Automatic Caption Generation for the Visually Impaired

Data Science Project Course Case Study by Vinya Ponugoti, December 2023

Prompt: Globally, around 2.2 billion people have some form of visual impairment. Visually impaired individuals oftentimes have limited access to information, technology, visual media, and education or employment opportunities. Improving online accessibility is crucial for these individuals as it can help them access digital content with fewer barriers and in turn have greater access to information. Integrating machine learning with accessibility features through image classification and converting what vision impaired people see to a format they can consume has been a goal for researchers. You have been hired as the data scientist for a visual impairment research group to explore how machine learning techniques can be used to enhance online accessibility through automatic caption generation for images. Automatic caption generation provides textual descriptions which can help individuals with visual impairments to access and understand the content of images on the internet, making web content more inclusive and accessible. Through the advancements of technology and the field of machine learning, there has been progress with the resources available to help the visually impaired but many of those resources also come with limitations, how can machine learning be used to help the visually impaired and are there limitations to the current models or data available?

Data: The data you will be using for this deliverable consists of University of Illinois' Flickr30k dataset, which has already been acquired for you and can be accessed through the google drive link here. The images that you will use to train your model are in that drive under *flickr30k-images.tar.gz* and can be accessed once downloaded. The captions that you will use to train the model are in the github inside of the materials folder here, it consists of the set of image names with each of their five corresponding captions. The well-lit set and the low-contrast set that you will generate captions for are in the github materials folder as well.

Deliverable: Create a caption generation model that is trained with image and caption data from the Flickr30k dataset to generate captions for a set of well-lit images and a separate set of low-contrast images. Using your own judgment, determine whether more accurate captions were created for the well-lit or low-contrast images. Pay attention to any limitations within your model. Create a slideshow illustrating your analysis process and conclusions. All your code should be in github in the appropriate format specified by the rubric. The datasets you will use as well as all other materials can be found in this github.