Andrew Chung
Project Proposal

Problem

The following problem description has been adapted from the introduction to the FashionIQ challenge at the following link.

Interactive image search has been studied to a certain depth in academia — however, there is still room for improvement. This especially applies when it comes to methods for connecting visual representations to the semantic information present in human language.

A potential application of being able to establish these connections between images and natural language would be a digital fashion assistant that finds specific apparel for you, based on your description of what you want. The objective of this project is just that — to create a chatbot model that will search for and retrieve images of apparel based on what the user describes in plain English. It is the same as the objective of the FashionIQ challenge, which was held the past couple of years.

Data

The data to be used in this project is the FashionIQ dataset, which has been tailor-made for this specific purpose of image retrieval via natural language feedback. The links for all image data are here, and the original GitHub repo with the labels is here.

Machine Learning Approach

This is essentially a recommendation system that depends on both computer vision (CV) and natural language processing (NLP) techniques. The final product — which can be a combination of multiple models — should take an input combination of ann initial image and relative captions (i.e., a short description of how the "target" image looks like *in comparison to* the initial image), and output the top *N* images that fit the description in the relative captions.

I will attempt to use a deep learning approach for both CV and NLP tasks, as that seems to be the general approach taken in most relevant papers to this project. It remains to be seen what model architectures will be best fit for the various aspects of this task.

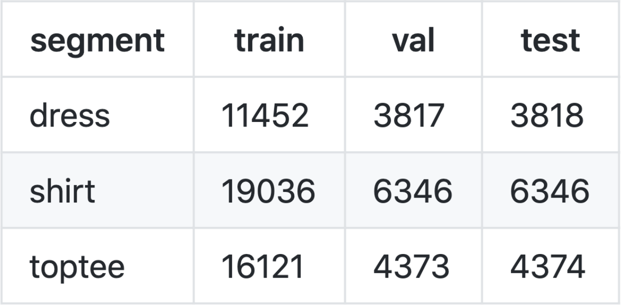
In the official FashionIQ competition, the quantitative metric used to measure performance of the recommender system was recall @ 10 and 50. Indeed, every sample in the test set also consists of a specific target image to which the captions refer — a performant final product would recommend the correct target image within the top *N* images for a small value of *N*.

Final Deliverable

For the final deliverable, I plan on capitalizing on my software engineering experience in both front- and back-end development to create a web-based chatbot that users can actually interact with and receive apparel recommendations from.

Computational Resources

The number of images in each partition is as follows:



segment	train	val	test
dress	11452	3817	3818
shirt	19036	6346	6346
toptee	16121	4373	4374

This is a total of 75,683 images, each fairly low resolution (around 10kb each). This equates to approximately 740mb for the size of just the images. The total size is not considerably different when considering the relative captions (i.e., the labels / annotations for this dataset).

Since the task does not require more than 12GB to 16GB of GPU memory, it seems that computation will not be an issue given the resources available on Paperspace.

Other resources

FashionIQ 2020, 2nd place repo: https://github.com/nashory/FashionIQChallenge2020